University of Plymouth

PEARL **https://pearl.plymouth.ac.uk**

04 University of Plymouth Research Theses 01 Research Theses Main Collection

2020

Detecting Energy Theft and Anomalous Power Usage in Smart Meter Data

Hock, Denis

http://hdl.handle.net/10026.1/16806

http://dx.doi.org/10.24382/855
University of Plymouth

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the author's prior consent.

UNIVERSITY OF
PLYMOUTH

DETECTING ENERGY THEFT AND ANOMALOUS POWER USAGE IN SMART METER DATA

by

DENIS HOCK

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Computing, Electronics and Mathematics

May 2020

ACKNOWLEDGEMENTS

This work was supported by the German Ministry of Economics and Technology under the Central Innovation Programme for SMEs (ZIM) Grant No. ZF4131801MS5 (ADMIN) and No. ZF4131804HB8 (iSMR).

Many people have contributed either directly or indirectly to this study. I cannot express enough thanks to my Director of Studies Prof. Dr. Martin Kappes for his comment and advices as well as assistance and guidance with this thesis. I offer my sincere appreciation for the learning opportunities, given by my supervisors Prof. Dr. Bogdan Ghita and Prof. Dr. Matthias Wagner.

My completion of this project could not have been accomplished without the support of my colleagues, Johannes Bouche, Manuel Grob, Dr. Robin Müller-Bady and Lukas Atkinson: my deepest gratitude.

I would also like to acknowledge here my gratitude for my co-authors M. Toulouse, H. Le, C. Phung from VGU.

Last but not least, I would like to thank my parents and friends for their continued support and encouragement.

Author's Declaration

At no time during the registration for the degree of *Doctor of Philosophy* has the author been registered for any other University award without prior agreement of the Doctoral College Quality Sub-Committee.

Work submitted for this research degree at *the University of Plymouth* has not formed part of any other degree either at *the University of Plymouth* or at another establishment.

This study was financed with the aid of a studentship from the *Frankfurt am Main Doctoral Node* and carried out in collaboration with *Frankfurt University of Applied Sciences*.

Word count for the main body of this thesis: 26,200

Signed: 

Date: 15th January 2021

Publications:

Bouché, J., Hock, D., & Kappes, M. (2016). On the performance of anomaly detection systems uncovering traffic mimicking covert channels. In *Proceedings of the 11th international network conference (inc)* (pp. 19-24).

Hock, D., Kappes, M., & Ghita, B. V. (2016). A pre-clustering method to improve anomaly detection. In *Proceedings of the 13th international joint conference on e-business and telecommunications (secrypt)* (pp. 391-396).

Toulouse, M., Le, H., Phung, C. V., & Hock, D. (2016). Robust consensus-based network intrusion detection in presence of Byzantine attacks. In *Proceedings of the 7th symposium on information and communication technology (soict)* (pp. 278-285).

Hock, D., Kappes, M., & Ghita, B. (2018). Non-intrusive appliance load monitoring using genetic algorithms. In *Proceedings of the 3rd international conference on renewable energy and smart grid (icresg)* (pp. 1-7).

Hock, D., & Kappes, M. (2018). Using the entropy for typical load curve classification. In *Proceedings of the 7th international conference on smart grid and clean energy technologies (icsgce)* (pp. 58-64).

Hock, D., Kappes, M., & Ghita, B (2020). Entropy-based metrics for occupancy detection using energy demand. *Entropy*, 22(7), 731.

Hock, D., & Kappes, M. (2020). A survey on the applications of energy demand. (Submitted to Elsevier RSER).

Hock, D., Kappes, M., & Ghita, B. (2020). Using multiple data sources to detect manipulated electricity meter by an entropy-inspired metric. *Sustainable Energy, Grids and Networks*, 21, 100290.

DENIS HOCK

DETECTING ENERGY THEFT AND ANOMALOUS POWER USAGE IN SMART METER DATA

Abstract. The success of renewable energy usage is fuelling the power grids most significant transformation seen in decades, from a centrally controlled electricity supply towards an intelligent, decentralized infrastructure. However, as power grid components become more connected, they also become more vulnerable to cyber attacks, fraud, and software failures. Many recent developments focus on cyber-physical security, such as physical tampering detection, as well as traditional information security solutions, such as encryption, which cannot cover the entire challenge of cyber threats, as digital electricity meters can be vulnerable to software flaws and hardware malfunctions.

With the digitalization of electricity meters, many previously solved security problems, such as electricity theft, are reintroduced as IT related challenges which require modern detection schemes based on data analysis, machine learning and forecasting. The rapid advancements in statistical methods, akin to machine learning techniques, resulted in a boosted interest towards concepts to model, forecast or extract load information, as provided by a smart meter, and detect tampering early on. Anomaly Detection Systems discovers tampering methods by analysing statistical deviations from a defined normal behaviour and is commonly accepted as an appropriate technique to uncover yet unknown patterns of misuse. This work proposes anomaly detection approaches, using the power measurements, for the early detection of tampered with electricity meters. Algorithms based on time series prediction and probabilistic models with detection rates above 90% were implemented and evaluated using various parameters. The contributions include the assessment of different dimensions of available data, introduction of metrics and aggregation methods to optimize the detection of specific pattern, and examination of sophisticated threads such as mimicking behaviour. The work contributes to the understanding of significant characteristics and normal behaviour of electric load data as well as evidence for tampering and especially energy theft.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This work proposes a comprehensive framework based on intelligent methods, for the early detection of various threats, especially energy theft, in the power grid. An emphasis is on anomaly detection, i.e. by detecting deviations from the expected behaviour. More specifically, on the analysis of electrical load curves that are generated by the power (W) data of intelligent meters in residential households.

To provide a basis for modelling the expected behaviour, the thesis analyses consumption curves and their characteristics. Based on these characteristics, metrics for the analysis and prediction of consumption curves, which are particularly suitable for anomaly detection, are designed and evaluated. Using these metrics, anomaly detection methods are investigated with regards to possible data sources, e.g. historical data or spatially close, identical devices. Last but not least, stealthy

energy theft and the limitations of anomaly detection systems, based on an evaluation with two different anomaly detection methods, are introduced.

The scientific contributions of the work are, in addition to a comprehensive literature survey and data analysis, metrics for the detection of human activity, exemplary methods for the normalization of metrics and methods for optimizing the parameters with regard to energy theft. This thesis also provides an anomaly detection method based on Holt Winters prediction with detection rate over 90% and the comparison to other anomaly detection methods based on Naive Bayes and time series decomposition using real world measurements. Furthermore, the thesis provides a performance comparison with advantages and disadvantages of different data sources for anomaly detection and a performance comparison of energy theft methods for different anomaly detection systems with respective methods to limit the damage caused by targeted mimicking attacks on anomaly detection systems.

1.1 Aims and Objectives

The high resolution electricity data at residential level, as provided by modern smart meters, can provide a better understanding of energy usage for both electricity producers and consumers. Many European coun-

tries launched a massive expansion of renewable energies and switched to intelligent power meters. These devices enable energy suppliers to react fast on variable energy demand and the volatile renewable energy. The advancing decentralization, together with the requirements to control and monitor devices in low-voltage networks, resulted in a power grid that no longer exclusively transmits energy but also data, and hence serves as a communication channel. This, so called, smart grid differs greatly from conventional power grids. As a next-generation electrical network, it features a two-way communication of sensors and actuators distributed in several logically separated networks, where large quantities of fine-granular information are collected and analysed. The high resolution electricity data at residential level, as provided by modern smart meters, can provide a better understanding of energy usage for both electricity producers and consumers.

Smart grids are expected to be resistant against disturbance or outages, which is challenging due to the large-scale use of decentralized volatile and renewable energy, such as wind and photovoltaic energy. Solving the challenge of accurate energy demand prediction and associated tasks such as the prediction of certain consumer profiles, dynamic energy pricing for particular consumer groups or monitoring of individual appliances is essential to enable optimal demand response. Further-

more, it is of paramount importance to protect the smart grid, as part of the critical infrastructure, from cyber attacks, fraud and software failures.

Over the last decades researchers addressed different aspects of energy demand in a number of research areas – to name only a few: Non-Intrusive Appliance Load Monitoring (NIALM), energy forecast, residential energy demand modelling and Typical Load Classification (TLC). These approaches have in common, that they aim to structure and organise the measured data in a way that allows to monitor aspects such as safety, security or efficiency of the power grid or parts of the power infrastructure. Many of the publications, of aforementioned research areas date back to the early 80s and are reintroduced in current smart grid research. Since high resolution data on individual households was not available in the traditional power grid, the scope and objectives of many early publications differ, but characteristic features of energy demand are still valid and can be applied to modern approaches.

The scope, or generally the usage of smart grid data in research, can be outlined by the wide range of available surveys. Many early studies related to the power grid and energy usage, e.g. Bohi and Zimmerman (1984), focus on the development of energy demand as well as corresponding modelling methods for large areas in the long-term. Chang,

Leung, Wu, and Yuan (2003) analysed the consumption and production of traditional and renewable energy in China. Connolly, Lund, Mathiesen, and Leahy (2010) reviewed computer aided tools for energy management. Banos et al. (2011) showcased optimization methods applied to renewable energy. Short-term forecasting also offers a wide range of literature, e.g. Suganthi and Samuel (2012) reviewed forecasting methods, categorized by the prediction method used. Foucquier, Robert, Suard, Stéphan, and Jay (2013) and Zhao and Magoulès (2012) reviewed methods to predict building energy consumption. Dutta and Mitra (2017) reviewed literature concerning energy forecast for dynamic pricing of electricity. Many recent approaches often focus on methods to model residential households, which is sometimes called appliance level load profile generation. Grandjean, Adnot, and Binet (2012) pointed out that recent load profile generation approaches can be categorized in time of use based or probabilistic models, while Swan and Ugursal (2009) introduced the terminology of bottom-up and top-down modelling techniques.

Rather than using household characteristics and socio-economic factors as input for forecasting, the observed load curves can instead be used to indicate those characteristics, which go far beyond the traditional classification of industrial and residential consumer. Energy provider

can use these information for Demand Side Management (DSM), by introducing differentiated and personalized tariffs according to the consumers habits or to monitor a particular customer base to detect deviations from the expected behaviour. Some authors such as Yaseen and Ghita (2017) even suggest to perform DSM by using customer profiles to compute an optimal schedule for intelligent appliances. However, the large quantity of information that can be drawn from energy demand also raise privacy concerns since, depending on the time granularity, personal habits and sensitive information about the household might leak. Anderson, Lin, Newing, Bahaj, and James (2017) listed a number of approaches to correlate household information to residential energy demand. Y. Wang et al. (2015) and S. L. Yang and Shen (2013) reviewed clustering methods, which have been on used on energy demand. Zoha, Gluhak, Imran, and Rajasegarar (2012) classified recent NIALM methods in schemes using the transition or the steady phase of appliances.

Other than that, the smart grid has a rich palette of topics related to communication and security as several surveys analysing the topics of smart grid security show. Anu, Agrawal, Seay, and Bhattacharya (2015); W. Wang and Lu (2013); Yan, Qian, Sharif, and Tipper (2012) summarized research on security requirements in the smart grid, such as countermeasures for network vulnerabilities, secure communication pro-

tocols and innovative smart grid architectures. Delgado-Gomes, Martins, Lima, and Borza (2015); J. Liu, Xiao, Li, Liang, and Chen (2012); McDaniel and McLaughlin (2009) presented and overview of relevant cyber security and privacy issues. Ericsson (2010) highlights the security of access points in a substation.

After this brief outline of research topics related to energy demand and smart meter data, a detailed investigation of the smart grid and anomaly detection follows in the next chapter. In the further thesis the related work of each individual topic is examined in the corresponding chapter.

The aim of this thesis is the concept, design and evaluation of an anomaly detection system to unveil energy theft early on, using power measurements. The thesis proposes solutions for modelling the expected behaviour of electric load curves, designs and evaluates metrics to detect anomalous power usage in residential load data and introduces a method to detect energy theft using anomaly detection based on time series prediction. In order to reach the aim of this thesis the objectives are to find evident security threats and their characteristics in energy demand. To find significant characteristics of electric load curves that can describe normal behaviour of residential households. Furthermore, to structure and organise this data from the available data sources into a

statistical model, so that anomalous data can be unveiled and to define the performance and limitations of anomaly detection. This works area of interest is centred around the data provided by residential smart meters in low-voltage areas. The scope includes, but is not limited to, the characteristics and normal behaviour of energy demand, statistical metrics to detect energy theft, an evaluation of different data sources and the usage of multiple data sources, as well as a comparison to alternative anomaly detection methods.

This study focuses on the analysis, modelling and processing of measurement data from a statistical viewpoint, which means that many aspects of electrical engineering, e.g. the technical requirements to physically tamper with electricity meters, as well as aspects of information security, e.g. the feasibility to manipulate the encrypted communication of a smart meter, are not in the scope of this work.

Some anomaly detection approaches, such as AMIDS by McLaughlin, Holbert, Fawaz, Berthier, and Zonouz (2013), analyse sensor data of smart meters (e.g. disconnection alerts, physical tampering alerts), but this work exclusively examines measurements and evaluate metrics and algorithms applied to the consumption data. Furthermore, in contrast to some works which use voltage or reactive power to better distinguish devices, this work solely uses real power (W) values.

The currently popular topic, which is related to the manipulation of smart meters, is false data injection to execute a Byzantine attack, as described by Lamport, Shostak, and Pease (1982). However, Byzantine attacks aim falsify a grid-wide state, e.g. to destabilize the mains frequency, which is not discussed in detail.

Inherent limitations of anomaly detection systems, such as poisoning the model during learning phase, a denial of service by flooding with false alerts or evading the detection by reverse engineering the statistical model are not central to this study. These issues are already well researched, e.g. by Corona, Giacinto, and Roli (2013), and apply to the smart grid context as well as any other context.

With the digitalization of electricity meters many previously solved security problems, such as electricity theft, are reintroduced as IT related challenges which require modern detection schemes based on data analysis, machine learning and forecasting. The aim of this study is to demonstrate anomaly detection approaches for the early detection of tampered with electricity meters.

Raw load curves are difficult to compare because the aggregated and overlaying patterns of several components can drastically change and pollute the appearance of a load curve. To clearly distinguish legitimate and anomalous data, it is crucial to filter any unnecessary information.

Throughout this work several metrics, which can characterise load curves well enough to distinguish a residential households normal behaviour from energy theft and significantly increase the detection rate of energy theft, are introduced.

In the course of this work, an emphasis is on following research questions:

RQ1 What are the most significant security threats in the smart grid?

RQ2 What are the defining characteristics of electric load curves?

RQ3 Is it possible to extract human activity from electric load curves?

RQ4 Is it possible to use different data sources as expected behaviour?

RQ5 Do multiple data sources improve the detection rate?

RQ6 Is it possible to detect stealthy manipulation attempts?

1.2 Thesis Structure

Chapter 2: *Overview of the Smart Grid* covers an introduction on the smart grid, with special regard to security aspects, including a threat taxonomy and the motivation to use next-generation anomaly detection schemes. Furthermore, background information about anomaly detection as well as evaluation methods used throughout this work

are provided. **RQ1** is investigated by introducing the architecture and fundamental knowledge of smart grids as well as the relationship to anomaly detection.

Chapter 3: *Analysis of Energy Demand Characteristics* aims to provide to the foundations of load curves with a comprehensive literature review on electricity data. The chapter first introduces the concept and mechanics of appliance load shapes and an analysis of load curve characteristics. **RQ2** is investigated by presenting and evaluating the unambiguity of a load curve using the entropy as a metric. The chapter demonstrates that the human activity is one of the characteristics which can be used to represent the normal behaviour of load curves.

Chapter 4: *Extracting the Human Activity* presents a concept to extract the human activity to investigate **RQ3**. Two methods as well as an analysis of the corresponding parameters are introduced with an evaluation of the accuracy in comparison to other statistical methods.

Chapter 5: *Normalized Characteristics of Energy Demand* provides an overview of common metrics used in smart grid applications and introduces normalized metrics to compare different data sources,

such as the smart meters of several households with each other. The chapter evaluates the quality of three features to detect energy theft and investigates **RQ4**.

Chapter 6: *Anomaly Detection with Multiple Dimensions* introduces the advantage of using several data sources and especially a method to remove the daily pattern from multiple sources while preserving outliers which represent energy theft. The chapter investigates **RQ5**. The contributions are an evaluation of different data sources as well as the comparison of the suggested method to alternative anomaly detection methods.

Chapter 7: *Analysis of Stealthy Energy Theft* showcases of stealthy energy theft methods, which are only feasible with a digital manipulation. The chapter introduces different concepts to mimic the expected behaviour according to the anomaly detection model while highlighting the maximum amount of stolen energy and then investigates **RQ6** by evaluating the efficiency of different anomaly detection methods.

Throughout the thesis, the results of the authors peer-reviewed articles, which are also mentioned at the beginning of each chapter, are used as follows:

Bouché, J., Hock, D., & Kappes, M. (2016). On the performance of anomaly detection systems uncovering traffic mimicking covert channels. In *Proceedings of the 11th international network conference (inc)* (pp. 19-24). introduces a concept for data falsification techniques in presence of Anomaly Detection Systems together with an analysis of the time complexity and limitations of sophisticated mimicry attacks. Is used in *Chapter: Analysis of Stealthy Energy Theft.*

Toulouse, M., Le, H., Phung, C. V., & Hock, D. (2016). Robust consensus-based network intrusion detection in presence of Byzantine attacks. In *Proceedings of the 7th symposium on information and communication technology (soict)* (pp. 278-285)., which presents algorithms, inspired by swarm intelligence, to assess the trustability of network participants and mitigate false data injection approaches. The concept of Byzantine attacks is briefly mentioned in some chapters.

Hock, D., Kappes, M., & Ghita, B. (2018). Non-intrusive appliance load monitoring using genetic algorithms. In *Proceedings of the 3rd international conference on renewable energy and smart grid (icresg)* (pp. 1-7)., proposes a genetic algorithm to extract the load curve of individual appliances from aggregated

load. While the concept of non-intrusive appliance load monitoring is only briefly mentioned, some fundamental knowledge mentioned in this paper is used in *Chapter: Analysis of Energy Demand Characteristics.*

Hock, D., Kappes, M., & Ghita, B. V. (2016). A pre-clustering method to improve anomaly detection. In *Proceedings of the 13th international joint conference on e-business and telecommunications (secrypt)* (pp. 391-396). which applies machine learning techniques as advanced data preprocessing to improve Anomaly Detection Systems - especially methods based on time series analysis. *Chapter: Anomaly Detection with Multiple Dimensions* contains some ideas of this publication.

Hock, D., & Kappes, M. (2018). Using the entropy for typical load curve classification. In *Proceedings of the 7th international conference on smart grid and clean energy technologies (icsgce)* (pp. 58-64). evaluates the entropy as similarity measure for the comparison of household load curves and presents an prototypical method to cluster load curves. *Chapter: Analysis of Energy Demand Characteristics* is based on the concepts of this article.

Hock, D., & Kappes, M. (2020). A survey on the applications of energy demand. (Submitted to Elsevier RSER). summarises a comprehensive literature review on state of the art techniques

used to analyse, model and monitor energy load curves. The literature of this survey is used throughout this thesis.

Hock, D., Kappes, M., & Ghita, B (2020). Entropy-based metrics for occupancy detection using energy demand. *Entropy*, 22(7), 731. presents the entropy as metric to detect human activity on individual household load curves. *Chapter: Extracting the Human Activity* is largely based on the results of this publication.

Hock, D., Kappes, M., & Ghita, B. (2020). Using multiple data sources to detect manipulated electricity meter by an entropy-inspired metric. *Sustainable Energy, Grids and Networks*, 21, 100290. shows a concept to utilize smart meter data of multiple sources to detect energy theft. Is used in *Chapter: Normalized Characteristics of Energy Demand* and *Chapter: Anomaly Detection with Multiple Dimensions.*

Chapter 2

Overview of the Smart Grid

2.1 Introduction

In broad outline, this chapter introduces the smart grid and fundamental knowledge of smart grid security. It will highlight background information needed to understand smart meter data and energy theft, including a presentation of the smart grid architecture and information security considerations, related research and an introduction of basic anomaly detection knowledge. Furthermore, evaluation methods, such as the different types of alerts and metrics used throughout this work are introduced. The chapter was designed to investigate RQ1: 'What are the most significant security threats in the smart grid?'. While a complete review of the last 35 years of research is out of the scope of this work, the studies objective is to introduce the necessary prerequisites to understand the following chapters with consideration to earlier approaches

and cover a wide range of generally accepted schemes and frequently used terms. Parts of this chapter are based on the article 'Hock, D., & Kappes, M. (2020). A survey on the applications of energy demand. (Submitted to Elsevier RSER).'. Note that, the related work corresponding to topics of the individual chapters can be found at the beginning of each individual chapter.

2.2 The Smart Grid

The massive expansion of distributed, renewable energy sources, such as wind and photovoltaic energy places new demands on electric power grids. While low-voltage networks in traditional power grids are conventionally unmonitored, the decentralized nature of volatile and renewable energy, and the fact that a large number of these power generators produce within low-voltage networks, results in a strong need to monitor and maintain grid stability. This can only be achieved through coordinated energy production and consumption. The smart grid, which connects all components of the energy system, aims to optimize the reliability and efficiency of energy production and consumption utilizing a two-way communication.

Smart Grid Architecture

The smart grid combines energy generation, storage and consumption to automatically optimize the operation of its interconnected elements, and thus compensates for fluctuations in performance, e.g. due to fluctuating renewable energies. The smart grid is defined by a bidirectional flow of electricity and information. Not only energy, but also data is transported in a smart grid, so that operators receive information on energy production and consumption.

In a technical report of the National Institute of Standards and Technology (NIST), Von Dollen (2009) identified a number of interacting components in the smart grid, which are responsible for the power consumption schedule according to customer preferences, a two-way communication between consumer and energy supplier, the monitoring and control of the power grid infrastructure and detection of failures as well as the optimized energy production.

These components include *intelligent appliances*, capable of deciding when to consume power based on the customer preferences and *smart power meters*, featuring two-way communications between consumers and energy supplier. Furthermore, *smart substations*, which include monitoring and control of data such as the configuration of power line switches, circuit breakers and batteries, as well as *smart distribution*

with automated monitoring and analysis tools capable of detecting failures based on real-time data and *smart generators*, tasked with the optimization of voltage, frequency and power based on feedback from the grid.

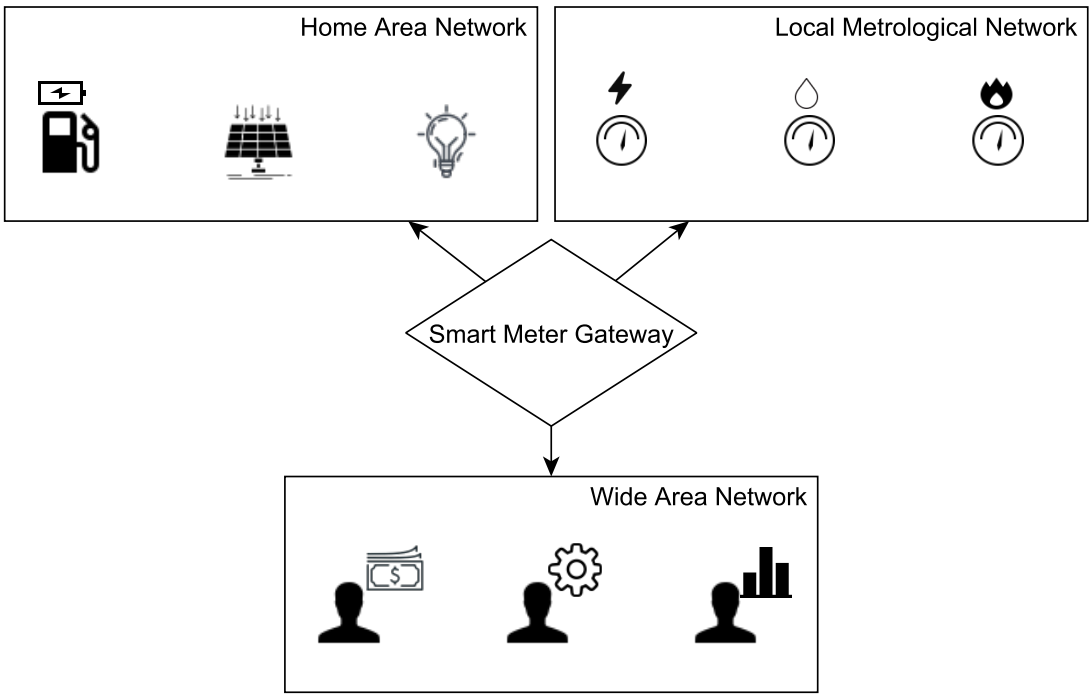


Figure 2.1: Example of a smart meter in a residential household.

This thesis focuses on the problems and challenges concerning the components in the low-voltage networks and especially focuses on smart meters. The author assumes an architecture with multiple gas, water, and electricity meters connected to a so called Smart Meter Gateway (SMGW), which collects the metering data and provides access to other authorized participants of the smart grid as outlined by Gungor et al.

(2012).

The SMGW in Figure 2.1, which depicts an example of such an architecture, aims to enable secure data transmission between the household and external market participants and functions as a firewall between three separate communication interfaces connected to the system operators and energy producers via Wide Area Network (WAN), the smart meters via Local Metrological Network (LMN) and end user or appliances in a household via Home Area Network (HAN).

The SMGW is typically installed in end consumers households and receives readings of the respective smart meters in the LMN. The LMN is used to communicate with the power meter or other meters. The data transmitted are typically energy consumption or production levels and measured parameters such as voltage, frequency or phase angle. This data is sent to external market participants using the WAN, which is an encrypted interface where authorized users can pull data.

The HAN can contain controllable consumers (e.g. tumble dryer) and decentralized energy generators (e.g. photovoltaic systems), also known as controllable local systems, which may be controlled via the SMGW by vendor specific protocols.

Using the fine-grained data obtained from the SMGW, many approaches aim to derive high level information from the fine-grained data.

Traditionally, the objectives range from monitoring approaches to the objective of creating reasonable dynamic pricing or optimize the energy usage through appliance monitoring. Many of those approaches have been summarized in reviews, e.g. Anderson et al. (2017) listed a number of approaches to correlate household information to residential energy demand, Zoha et al. (2012) classified recent NIALM methods and Y. Wang et al. (2015) reviewed clustering methods.

Compared with the traditional method of placing sensors on each individual end-use, the idea of NIALM, to disaggregate data acquired from a single point of measurement, is very cost efficient and convenient because energy suppliers are still not capable of deploying appliance level sensors in all households. Hart (1992) was the first to publish a method to distinguish appliances in aggregated load, which is today used for many other applications such as anomaly detection or clustering. Hart used a simple edge detection algorithm to find change-points and clustering to extract appliance patterns and build a state machine, modelling the different states of each appliance.

Figure 2.2 illustrates appliance on/off transitions with an artificial example, for devices such as lamps, with power (y-axis) over time (x-axis). The different states of the appliance load curve are labelled according to Hart as 'steady' for regions without a change of power and 'transition'

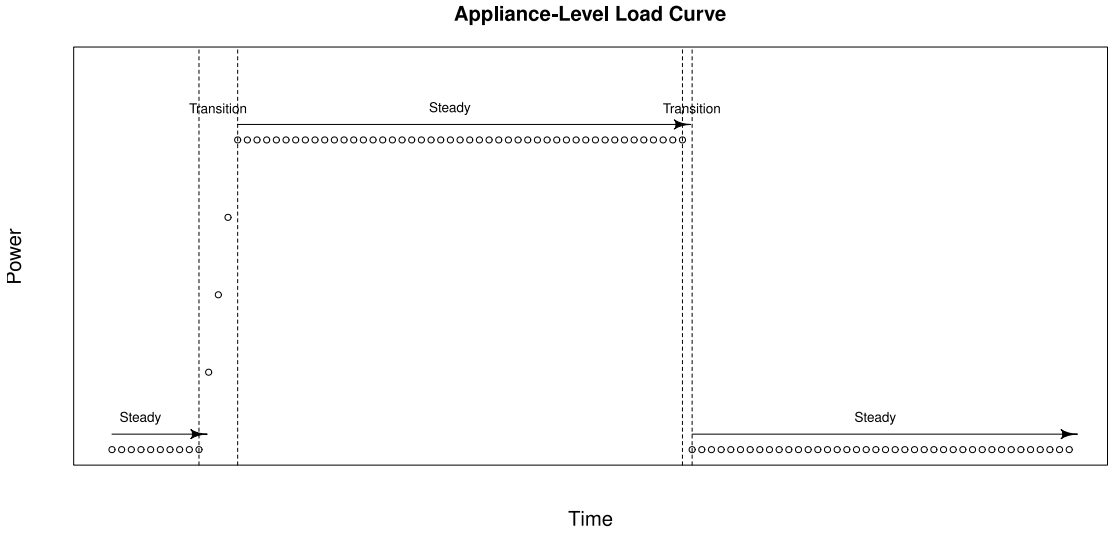


Figure 2.2: Steady and transition phase example.

for regions with changes above a certain level.

Many authors extended and improved Hart's method, e.g. Zeifman (2012) used a maximum likelihood classifier for energy disaggregation. Kim, Marwah, Arlitt, Lyon, and Han (2011) generated Hidden Markov models to detect appliances. Saitoh, Osaki, Konishi, and Sugahara (2010) improved Hart's approach using clustering methods. Baranski and Voss (2004) used optimization methods to reveal appliances and Ruzzelli, Nicolas, Schoofs, and O'Hare (2010) trained artificial neuronal networks to identify appliances.

Another popular research area, which requires extracted information is TLC, representing coherent groups of consumers, which can be used to improve the accuracy of load forecasting for 'typical' load profiles of

certain households. The objective of such profiles include differentiated and personalized tariffs according to the consumers energy demand, but such predefined consumers groups can also be utilized as a normal model for anomaly detection.

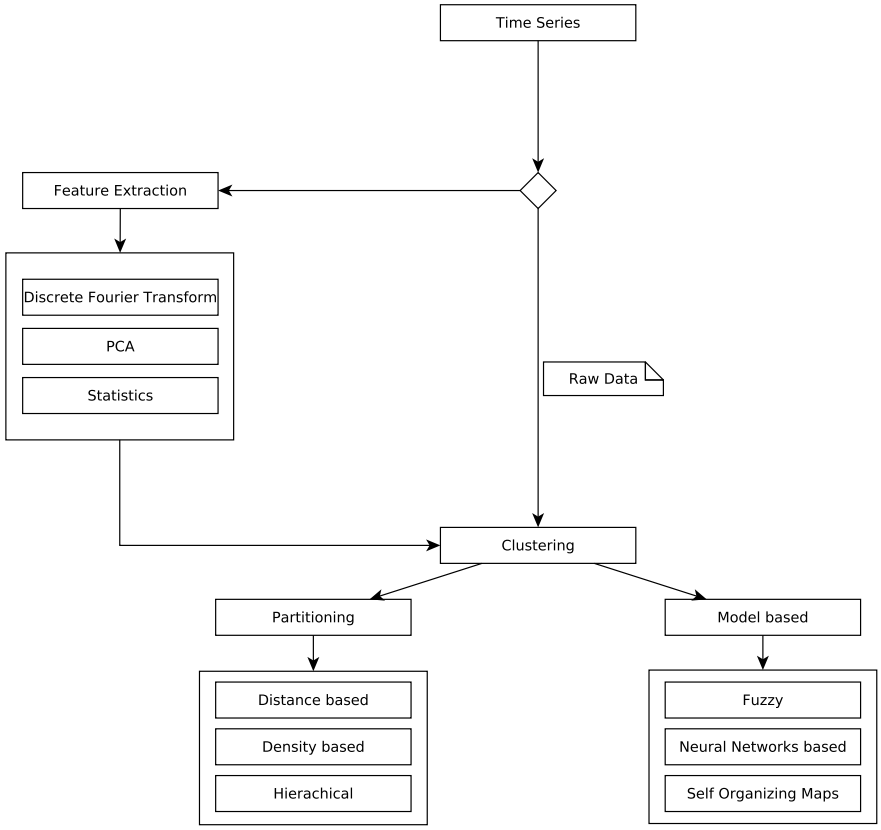


Figure 2.3: Time series clustering approaches.

Figure 2.3 presents a small taxonomy of features and clustering algorithms used to classify households. Since load curves are time series, most classification methods include a data preprocessing step to extract features.

Jota, Silva, and Jota (2011) clustered typical daily load curves with

hierarchical clustering by using the mean value of time windows. Xiao, Yang, Que, Li, and Gao (2014) used Haar wavelets to model electricity consumption data and then used k-means to obtain typical load curves. Chicco and Ilie (2009); Chicco, Ionel, and Porumb (2013); Chicco, Napoli, and Piglione (2006) used fuzzy k-means, electrical pattern ant colony clustering and support vector clustering to classify load curves and McLoughlin, Duffy, and Conlon (2013) uses Fourier transforms and Gaussian processes to characterise energy demand. Zakaria and Lo (2009) applied Principal Component Analysis (PCA) before fuzzy k-means clustering.

Another popular research topic in smart grids is forecast and prediction. This topic is related to anomaly detection as the predicted can be compared to the actual data in order to detect deviations. However, released studies range from the long-term annual prediction of national energy usage to the prediction of individual households. Forecasting energy demand offers a wide range of literature, which is summarized in several surveys. To name only a few, Suganthi and Samuel (2012) reviewed forecasting methods, categorized by the prediction method used. Swan and Ugursal (2009) distinguishes between bottom-up and top-down modelling techniques to predict energy demand. Foucquier et al. (2013) and Zhao and Magoulès (2012) reviewed methods to predict

building energy consumption. Dutta and Mitra (2017) reviewed literature concerning energy forecast for dynamic pricing of electricity.

Table 2.1: List of time series analysis methods.

Author	**Year**	Method
Hirst et al.	1977	Regression
Parti et al.	1980	CDA[1]
Aigner et al.	1984	CDA[1]
Cho et al.	1995	ARIMA[2]
Hunt et al.	2003	Trendanalysis
Labandeira et al.	2005	Regression
Tiedermann	2007	CDA[1]
Arroyo et al.	2007	Exponential Smoothing
Kavaklioglu et al.	2009	SVM[3]
Hong et al.	2009	SVM[3]

The applications of forecasting range from long-term planning and DSM to energy demand simulation and anomaly detection. Table 2.1 shows ten prediction methods sorted by their release date. The table shows a trend from long-term, large-scale prediction to short-term residential prediction. Due to the lack of fine-grained data and ubiquitous measurements in the low-voltage networks, many early approaches aimed to predict the demand in the long-term and large-scale or used supplementary data to draw conclusions on residential level. Hirst, Lin, and Cope (1977) initiated the first top-down model of annual scale for the energy usage of the USA. The model used demographic, economic, and technological factors in order to model residential energy usage with regression analysis. Tiedemann (2007) used a consumer survey

[1]Conditional Demand Analysis (CDA)
[2]Autoregressive Integrated Moving Average (ARIMA)
[3]Support Vector Machine (SVM)

together with billing and weather data to calculate annual energy consumption estimates for 14 end-uses. Labandeira, Labeaga Azcona, and Rodríguez Méndez (2006) used similar inputs for a regression model to model the Spanish residential energy consumption.

Parti and Parti (1980) initiated the research on another regression method, called CDA, which attempts a more accurate estimation by disaggregating demand into the end-use loads to estimate residential energy use. Aigner, Sorooshian, and Kerwin (1984) employed CDA with energy demand in a resolution of 15 Minutes, obtained from 130 households, to generate hourly appliance load profiles. These were used as the input for regression equations, one for each hour of the day, which predicted the daily energy demand. Hunt, Judge, and Ninomiya (2003) analysed underlying trends and seasonality of energy loads. Arroyo, San Roque, Maté, and Sarabia (2007) used exponential smoothing methods to predict energy demand. Cho, Hwang, and Chen (1995) applied ARIMA to forecast energy demand. A relatively new approach is Support Vector Regression to predict future demand, as presented by Kavaklioglu, Ceylan, Ozturk, and Canyurt (2009) and Hong (2009).

Table 2.2 shows machine learning approaches aiming to predict energy demand. As the release year shows, the topic of machine learning established later in contrast to other forecasting approaches.

Table 2.2: List of machine learning methods.

Author	**Year**	Method
Kalogirou et al.	2000	ANN[4]
Datta et al.	2000	ANN[4]
Tzafestas et al.	2001	GA[5]
Aydinalp et al.	2002	ANN[4]
Ozturk et al.	2004	GA[5]
Yang et al.	2005	ANN[4]
Sadeghi et al.	2011	GA[5]
Ghanbari et al.	2013	Ant Colony Opt
Liu et al.	2014	GA[5], SVM
Ribeiro et al.	2020	ANN[4], Bayesian Optimization

In several publications Ozturk, Canyurt, Hepbasli, and Utlu (2004) tested approaches using a GA to determine energy consumption in Turkey. Sadeghi, Zolfaghari, and Heydarizade (2011) proposed a 'Genetic Algorithm Electricity Demand Model' for forecasting. Tzafestas and Tzafestas (2001) aimed for short-term electric load forecasting by training an ANN with a GA. D. Liu, Niu, Wang, and Fan (2014) optimized a SVM using a GA. Ghanbari, Kazemi, Mehmanpazir, and Nakhostin (2013) used ant colony optimization to build an expert system which predicts energy demand.

J. Yang, Rivard, and Zmeureanu (2005) showcased an adaptive ANN, which uses a sliding window of energy demand and weather data, to forecast building energy consumption. Kalogirou and Bojic (2000) used an ANN to predict the energy demand of a building using properties such as the wall thickness. Aydinalp, Ugursal, and Fung (2002) sepa-

[4]Artificial Neural Network (ANN)
[5]Genetic Algorithm (GA)

rately modelled the energy demand of appliances, lighting and cooling with an ANN. Datta, Tassou, and Marriott (2000) aimed to predict the energy consumption in a supermarket with an ANN. Trierweiler Ribeiro, Guilherme Sauer, Fraccanabbia, Cocco Mariani, and dos Santos Coelho (2020) used Bayesian optimization on an Echo State Network for the prediction.

Smart Grid Security Aims and Threats

As part of the critical infrastructure hedging the smart grid against failures and errors is essential. Even robust components can be subject to malfunctions and failures, which must not jeopardize the operation of the entire network. Due to the novelty of smart grids, efficient monitoring systems, able highlight anomalies and identify critical events are yet unexplored.

Aside from safety aspects, the smart grids also introduce new security risks due to the lack of physical separation between interconnected controllable local systems and network management, which opens vulnerabilities to e.g. malicious software, billing fraud or network interferences.

Smart grids are historically not designed with internet security in mind, as mentioned by Jain and Tripathi (2013), but security flaws can result in customer information leakage and a cascade of inadvertent or

deliberate failures, such as a massive blackout and destruction of infrastructures as introduced by Metke and Ekl (2010). Large-scale industrial control systems, which are used for smart grid communication, often used closed networks and proprietary industrial communication protocols such as Modbus, DNP3 or S7, but with time it has become more convenient and cost-effective to connect them to the Internet. Hence, smart grid security is a rapidly growing research area.

In their NIST special publication Ross, McEvilley, and Oren (2018) defined the key concepts of security as confidentiality, integrity, and availability, whereas integrity and availability extends to objectives of non-security disciplines such as performance, reliability and safety. When applied to the smart grid, the goals of IT security can be interpreted as follows:

- Confidentiality: sensitive information should be protected against unauthorized access by third parties. The data collected from the smart meters, as well as the customers personal information should not be accessed by any unauthorized entities. Compromising the confidentiality could mean that a potential malicious user can forge a false identity to gain access to the SMGW (*spoofing*) or gain more rights than actually provided for his role by exploiting a vulnerability (*privilege escalation*).

- Integrity: data transmitted should be protected against changes by unauthorized persons. Any control data to HAN devices should be received and implemented correctly and completely. The electricity price and bill should not be manipulated by unauthorized third parties. Compromising the integrity could mean to affect other participants with corrupt data (*tampering*) or to gain advantage by denying certain activities (*repudiation*).

- Availability: a SMGW should be available to legitimate users whenever they access. Some specifications require, that connections are only established to previously defined recipients at defined times or after a wake-up request executed by the SMGW administrator. Compromising the availability could mean to influence the connection of the system (*denial of service*) for energy theft or to prevent the correct prediction of energy demand.

Many recent developments focus on cyber-physical security, such as physical tampering detection to protect the integrity, as well as traditional information security solutions, such as the encryption of certain communication channels to protect the confidentiality and integrity. However, information security cannot cover the entire challenge of cyber threats, as such digital meters can be vulnerable to software flaws

and hardware malfunctions. Illera and Vidal (2014), demonstrated that smart meters installed in Spain used strong symmetric encryption, but stored a static encryption key in a plain text file, which allowed adversaries to artificially manipulate and tamper with the data and measurements of a smart meters communication channel. Recently, Westerhof (2017) simulated the disastrous consequences of a coordinated cyber attack on photovoltaic systems, which may lead to a national power outage. In a study of Dabrowski, Ullrich, and Weippl (2017), IoT bot nets use common devices, connected to the Internet, to selectively increase and decrease power consumption which can lead to falling below the standard frequency and ultimately to power outages. For this reason, there is a high interest in utilizing the fine-grained data and recent advances in machine learning to detect data manipulation and anomalies.

Another recent challenge of smart grid security is the detection of energy theft and tampering of smart meters. Tampering methods can be split to intrusive methods inside the meter housing and non-intrusive methods outside the meter. Common intrusive methods include attaching electrically conductive object to pass current away from the measurement circuit, disconnecting the phase to interrupt the measurement or exchanging the phase connection to archive a negative measure-

ment. Non-intrusive measurements include the usage of strong magnets to temporary disable the power supply of the meter. In addition to the physical methods mentioned above, modern smart meters pose the additional danger of digital manipulation of the data. In contrast to a relatively easy to detect interruption, or even negative energy demand, data manipulation can be forged to be more stealthy and difficult to detect.

A particular active and challenging field related to tampering is false data injection, which is a (tampering) method where adversaries attempt to compromise the information exchange of meters. Many authors proposed false data injection attacks as theoretical concept, e.g. Y. Liu, Ning, and Reiter (2011) showed the potential to influence the mains frequency using remote controlled devices to cause power outages and L. Xie, Mo, and Sinopoli (2010) specified a falsification method to manipulate the electric market price information. Researchers, such as Esmalifalak, Nguyen, Zheng, and Han (2011); Giani et al. (2011); Pasqualetti, Carli, and Bullo (2011) introduced new false data injection scenarios, while researchers such as Bobba et al. (2010); Kosut, Jia, Thomas, and Tong (2010) developed methods to identify falsely injected data.

The terms 'data falsification' and 'false data injection' are coined by

mitigation techniques developed for cognitive radio networks and therefore, often associated with the Byzantine generals problem of Lamport et al. (1982), which covers the consensus agreement in presence of compromised sensors. In this work, the term data falsification is generally used to refer to the tampering of smart meter data, aiming for energy theft and billing fraud, and explicitly not referring to the manipulation of a grid-wide system state aiming to cause blackouts.

The authors Depuru, Wang, and Devabhaktuni (2011); Nagi, Yap, Tiong, Ahmed, and Mohammad (2008); Nizar, Dong, and Wang (2008) employed machine learning to classify consumption pattern and load profiles in order to detect electricity theft. Cárdenas, Amin, Schwartz, Dong, and Sastry (2012) developed a game theoretic scheme between adversary and billing company. Bandim et al. (2003) proposed a central observer to compare the total energy consumption with the reported consumption of individuals. Salinas, Li, and Li (2013) suggested a distributed algorithm to compute the trustworthiness of each participant. Spirić, Dočić, and Stanković (2015) detected energy theft by monitoring the energy consumption with XMR charts. In addition to the physical methods mentioned above, modern smart meters pose the additional danger of sophisticated digital manipulation methods, which are not covered in this work. With digital access to measurements and me-

tering information as well as knowledge on the detection method, the adversary could potentially aim for stealthy manipulation scenarios such as mimicry attacks. However, such methods are general limitations of anomaly detection systems and not specific to energy theft, as introduced by Urbina et al. (2016) and Bouché, Hock, and Kappes (2016).

2.3 Anomaly Detection

Unfortunately, many general-purpose monitoring approaches cannot be effectively adopted to monitor smart grid environments. Traditional intrusion detection systems, such as Snort, and Suricata, depend upon a vast amount of attack patterns and permanent maintenance which is unsuitable for smart grid environments and detection approaches which focus extensively on content data and measurements.

A potential method to unveil energy theft and tampering is anomaly detection. Anomalies are defined as deviations from the expected data, rather than by predefined malicious data. Anomaly detection systems report any deviation of the normal behaviour, and therefore also recognize unknown tampering patterns. Therefore, anomaly detection is particularly suitable to unveil tampered data, which is difficult to describe in the volatile and highly heterogeneous energy demand.

Since Denning (1987) published the initial paper on anomaly detec-

tion, a wide range of techniques based on statistics, machine learning and soft computing techniques have been released for various information security contexts. The emergence of sensors with processing and communication capabilities stimulated great interests in anomaly detection for the Internet of Things and devices related to the smart grid, as shown by surveys of Zhang, Meratnia, and Havinga (2010) and M. Xie, Han, Tian, and Parvin (2011). An often referenced approach are so called 'context-aware' anomaly detection methods, which take several data sources into consideration. Frolik, Abdelrahman, and Kandasamy (2001) used data fusion with fuzzy logic for the aggregation of quasi-redundant sensor data. Bettencourt, Hagberg, and Larkey (2007) evaluated context-aware methods to identify measurement errors relative to its neighbours. Shah, Desrosiers, and Sabourin (2015) used tensor factorization for his contextual anomaly detection approach. Furthermore, Braun et al. (2012) used the minimum covariance determinant to detect faults in photovoltaic arrays and Dienst, Schmidt, and Kühne (2013) consults change-point analysis to observe the condition of photovoltaic power plants. Andrysiak, Saganowski, and Kiedrowski (2017) presented a solution to detect energy theft with network traffic anomaly detection in critical smart metering infrastructure. Mookiah, Dean, and Eberle (2017) introduced a graph-based anomaly detection approach,

where vertices represent smart appliances and edges represent their usage, to detect anomalies in power usage. Furthermore, Raciti and Nadjm Tehrani (2013) designed smart meters embedded with anomaly detection to identify threats on cyber-physical systems.

Besides these general anomaly detection schemes, many approaches aim to solely monitor the measurements provided by smart meters to detect anomalies. E.g. McLaughlin et al. (2013) developed the anomaly detection scheme called AMIDS, based on network data and power measurements, in order to detect energy theft. AMIDS utilizes energy demand together with a NIALM database to label the amplitude changes in a time series and subsequently learns benign and illegitimate behaviour with the Naive Bayes algorithm. Rossi, Chren, Buhnova, and Pitner (2016) proposed to take collective and contextual anomalies into account to detect events such as over-voltages and under-voltages. Yip, Tan, Tan, Gan, and Wong (2018) presented an anomaly detection scheme that adopts linear programming to detect energy theft and reduce false positives by taking into consideration the impact of technical losses and measurement noise. Fengming et al. (2017) detected anomalies, such as short circuit faults, by comparing a time series of measurements reconstructed by a recurrent neural network with the original data. Zhou, Zou, Arghandeh, Gu, and Spanos (2018) aimed to detect outliers such

as communication failures and voltage disturbances by comparing multiple time series of voltage with a randomized block coordinate descent algorithm.

The next section introduces key concepts of anomaly detection and evaluation metrics used in this work. However, this work assumes the reader to be familiar with general statistics and only aims to give a short recap on the foundations necessary to understand this work. One can find a good introduction in standard references such as *Statistics* by Freedman, Pisani, and Purves (2007).

Models and Decision Process

Contrary to other techniques, anomaly detection does not require the existence of attack patterns. It detects statistical deviations from normal behavior instead, which results in the capability to detect otherwise undiscovered tampering approaches.

Overall, an anomaly detection system assumes benign data to fall in a certain range, and hence defines a threshold to split the normal and anomalous data, whereas the results depend upon the overlap of data. To make a decision, anomaly detection methods typically attribute a probability or anomaly score s to each observation indicating its abnormality. The binary result $r \in \{0, 1\}$, where 0 denotes normal and 1

anomalous, is computed using a threshold ε.

$$r = \begin{cases} 1 & \text{if } s \text{ is } > \varepsilon \\ 0 & \text{if } s \text{ is } \leq \varepsilon \end{cases} \tag{2.1}$$

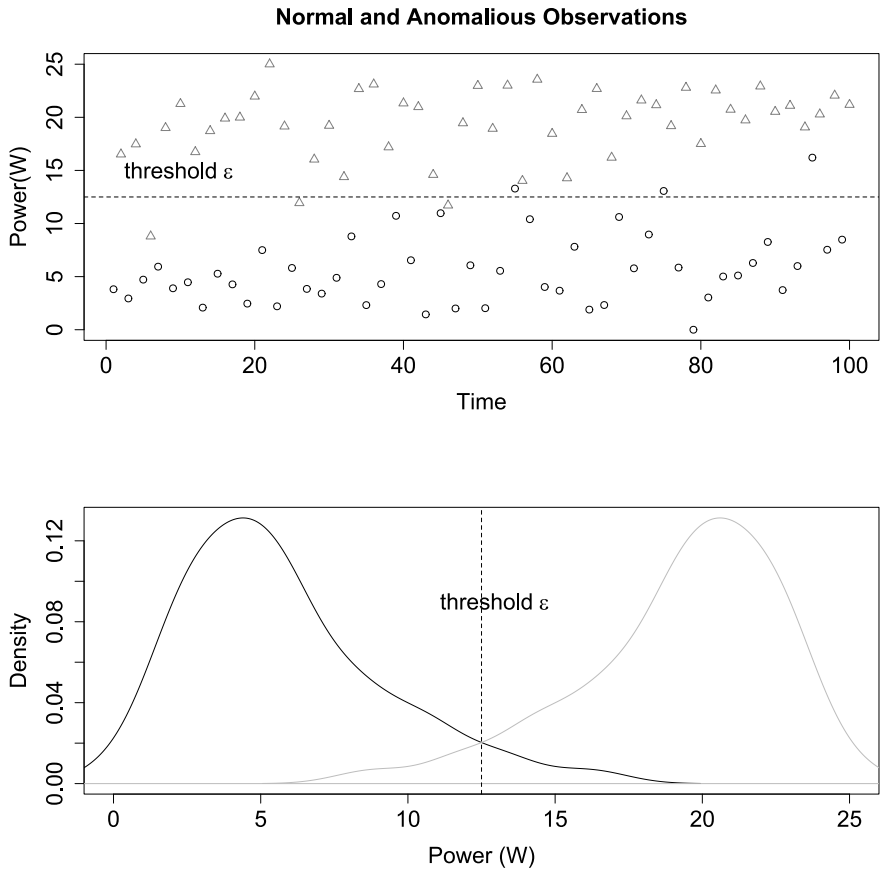


Figure 2.4: Visualisation of the classification problem.

Figure 2.4 illustrates the concept of classification based on thresholds with an artificial example. Note that, the distribution does not depict the distribution of real world power measurements or anomalous behaviour, instead the example aims to explain the general concept. For

simplicity, the power (W) is be used as anomaly score s in this example. The example includes two distinct set of data, benign and illegitimate data (points and triangles), with a total of $N = 100$ observations. The threshold cannot clearly divide both sets, due to the skewed distribution of points and triangles. In this work illegitimate activity defines intentionally or inadvertently harmful activities, such as energy theft or data falsification which is typically a subset of anomalous activity. The upper plot symbolises normal data (points) and illegitimate data (triangles), The dotted line is the threshold ε. If an observation was assigned a s above the threshold ε, but was actually legitimate behaviour (dot), it is called a False Positive (FP). If an event was assigned a s below ε, but was actually illegitimate behaviour (triangle), it is called a False Negative (FN). Correctly assigned results are called True Positive (TP) in case of anomalies and True Negative (TN) in case of legitimate behaviour. The bottom plot shows the density curves of the same example. The area where both sets overlap is a region where a threshold can not reliably separate normal and illegitimate data.

The above categories can be visualized in a so called confusion matrix, as shown in Table 2.3 and are typically used to evaluate the performance of anomaly detection.

Table 2.3: Statistical classification for the detection performance.

Confusion Matrix			
		Predicted	
		Legitimate	Illegitimate
Actual	Legitimate	TN	FP
	Illegitimate	FN	TP

Evaluation Metrics

The confusion matrix helps to calculate probabilities such as *sensitivity* (2.2), *specificity* (2.3) and *accuracy* (2.4) used to understand the performance of the detection.

$$Sensitivity = \frac{TP}{TP + FN} \tag{2.2}$$

The *sensitivity*, also known as true positive rate, is the probability to correctly identify all illegitimate activities. A low *sensitivity* implies many false negatives.

$$Specificity = \frac{TN}{TN + FP} \tag{2.3}$$

The *specificity*, also known as true negative rate, is the probability to correctly identify legitimate activities. A low *specificity* implies many false positives.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{2.4}$$

The *accuracy* is the total percentage of correctly identified activities. A 100% *accuracy* means that the actual activities are exactly as predicted by the anomaly detection system.

The performance of an anomaly detection model is often evaluated by contrasting sensitivity and specificity using the Receiver Operating Characteristic (ROC). One of the earliest references to the ROC curve, as an accuracy index for a statistical hypothesis to distinguish electric signals from noise, was provided by Peterson, Birdsall, and Fox (1954).

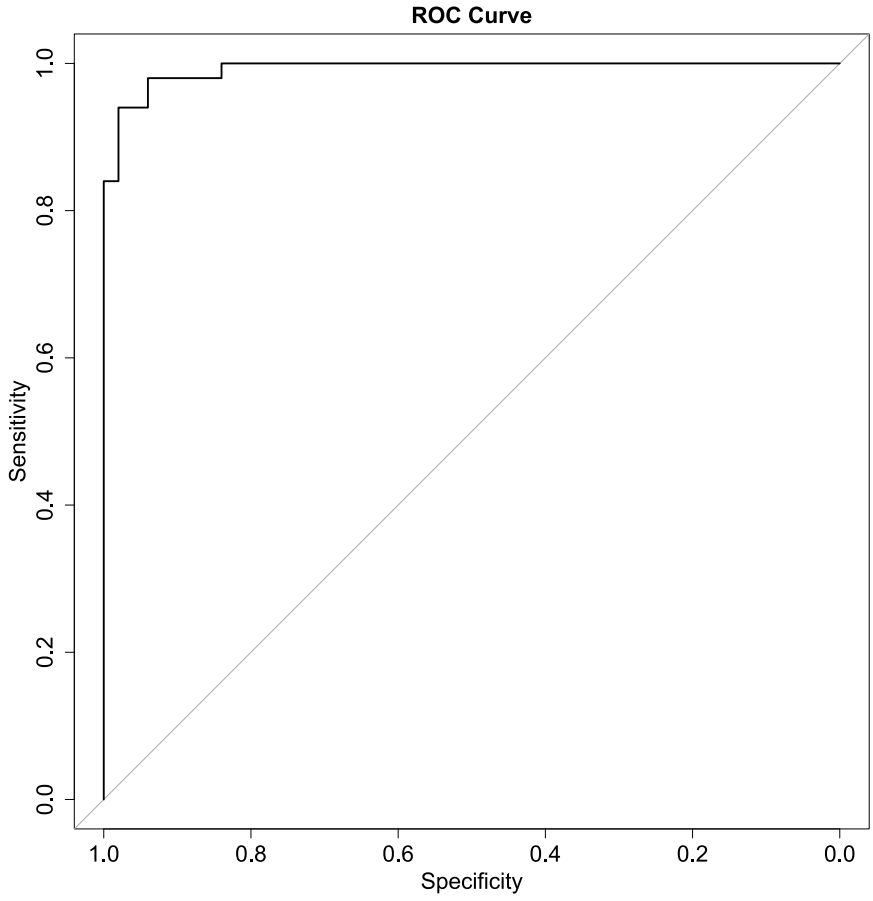


Figure 2.5: Visualisation of the ROC curve.

Figure 2.5 shows a ROC curve visualizing the example data of the previous Figure 2.4. A ROC curve shows the sensitivity (true positive rate) and the specificity (true negative rate) for any possible threshold to separate benign and anomalous values. The accuracy of a method can be defined by the Area Under the Curve (AUC). A good method can maximize both sensitivity and specificity, which results in a big area under the curve, while a random method results in a diagonal line and an AUC of 0.5. This work may refer to the AUC as detection accuracy.

2.4 Discussion

Here the smart grid architecture as well as threats and security aims of modern smart grid infrastructure were introduced. The first part of the chapter introduced many research topics related to the smart grid to give an overview of the applications of energy demand and smart grid data. Many forecast and clustering methods can be remodelled to detect anomalies, and hence the methods and metrics used in these state of the art works are important references. To investigate RQ1, the author briefly introduced the smart grid architecture and described possible threats on confidentiality, integrity and availability for the smart grid. The chapter motivated that measures of information security, such as encryption, are not sufficient to ensure the safety and security of smart me-

ters. Intrusive and non-intrusive tampering methods were distinguished in a brief introduction of energy theft methods. Data falsification and false data injection, which are related to Byzantine attacks, were delimited from the scope of this work. Furthermore, anomaly detection as well as related research in the smart grid context was introduced as solution to unveil threats such as energy theft, which is the focus of this work. Subsequently, the corresponding mathematical preliminaries have been introduced, including the decision process to create a binary result from a probability or anomaly score and typical metrics such as the AUC of a ROC curve to measure the accuracy of anomaly detection.

Chapter 3

Analysis of Energy Demand Characteristics

3.1 Introduction

Anomaly detection systems point out observations that do not correspond to the expected normal behaviour. For this reason, anomaly detection schemes must implement a statistical model of the normal behaviour and archive an divergence from this model in case of malicious scenarios. This chapter assesses the role of the normal behaviour as well as the difference to anomalous instances, by critically examining the typical energy demand load curve. The first section traces the development of publicly available datasets and explains in particular the question of available measurements and the extent of usage as well as resolution in related research. After an investigation of typical appliance load shapes

and appliance types, the issue of modelling techniques and parameters of bottom-up modelling methods such as appliance data and time of use data are discussed. Although these modelling techniques are used for simulation rather than forecast or anomaly detection, they show typical components of electric load curves, such as human activity. These characteristic components are utilized to clearly identify or distinguish residential energy demand which is, in the context of anomaly detection, necessary to distinguish illegitimate curves from normal ones. The following experiments show that certain time windows of a load curve, with unambiguous 'human activity', are better suited to differentiation than other time windows. In this chapter, RQ2: 'What are the defining characteristics of electric load curves?' is investigated. Note that, large parts of this chapter are based on the publication 'Hock, D., & Kappes, M. (2018). Using the entropy for typical load curve classification. In Proceedings of the 7th international conference on smart grid and clean energy technologies (icsgce) (pp. 58-64).'.

The efficient monitoring of low-voltage networks is one of the main arguments in favour of the introduction of smart meters. The basic function of a power meter is the measurement of the energy consumption, which is utilized for the billing. The meter computes the power by adding up the mains voltage multiplied with the current, drawn by all

active devices. Unlike the mechanical Ferrais counter, a smart meter can send digital readings at higher intervals to the meter operator. Typically, the measurements provided in the smart grid include the real (W), reactive (VAR), and apparent (VA) power as well as phase volt measurements (V), current (A) and mains frequency (Hz). These acronyms refer to the measurement units: according to the international unit system, volt is the unit for electrical voltage (U) and describes the amount of energy that is present in the individual electrons. Ampere is the unit for electrical current (I) and refers to the amount of electrons that flow through a line in a certain period of time. The electrical power (P) is the product of current times voltage and is given in watts.

Monitoring the current and voltage is critical for many applications such as the fault monitoring and the early detection of over-voltage or power line failures as proposed by Livani and Evrenosoglu (2013). Apart from that, the reactive and apparent power need to be closely monitored and counterbalanced by power grid operators to avoid unnecessary thermal line losses. Hart (1992) used the reactive power to identify appliances, since reactive power is caused by appliances with inductor. The mains frequency, which measures the balance of production and consumption, helps to prevent overproduction that potentially destroys equipment and underproduction that can ultimately lead to

blackouts.

From this perspective, the manipulation of each measurement can lead to safety critical situations. However, the primary goal of this thesis is to detect energy theft, and hence if the adversary tampered with energy demand values. Hence, this thesis focuses on the integrity of real power (W) measurements, which depicts the amount of work performed by a component. Manipulating power measurements can be seen as the foundation for many other sophisticated scenarios, such as a forged blackout and other scenarios requiring multiple corrupted electricity meters. Since collecting real world data is often time consuming and expensive, researchers released a number of clearly defined public datasets, with emphasis on power, monitoring household electricity and ambient parameters, enabling others to compare and evaluate their approaches against common benchmarks. These datasets can be independently validated and compared for reproducible scientific results and to prove the validity of a method in real life settings.

Overview of Public Datasets

Public energy demand datasets have been used as a evaluation methodology for many previously introduced smart grid topics, including but not limited to NIALM with the Electricity Consumption & Occupancy

Table 3.1: List of public energy demand datasets.

Name	Year	Level	Duration
Tracebase	2011	Appliance	12 month
REDD	2011	Whole-house, Appliance	1 month
IHEPC	2012	Circuit	47 month
Smart*	2012	Circuit, Appliance	3 month
AMPd2	2012	Circuit	24 month
ECO	2012	Whole-house, Appliance	8 month
iAWE	2013	Whole-house, Appliance	3 month
Greend	2014	Whole-house, Appliance	12 month
REFIT	2016	Whole-house, Appliance	24 month
UK Dale	2017	Whole-house, Appliance	48 month

(ECO) dataset by Makonin (2016), TLC with the Greend dataset by Andrade, Sampaio, Viterbo, Silva, and Boscarioli (2016) and occupancy detection with the Smart* dataset by D. Chen, Barker, Subbaswamy, Irwin, and Shenoy (2013). For a quick overview, Table 3.1 shows ten datasets and their corresponding properties in chronological order, whereas missing references do not imply that a particular dataset is unimportant. The datasets are categorized by their aggregation level and duration. In contrast to other public datasets, such as network traffic with many trending protocols and applications, energy demand data cannot become outdated. The table still includes the release year as it could influence the relevance and availability of the dataset.

The aggregation levels (see Figure 3.1), which show the physical connections of the power meter are plug (individual devices), circuit (typically one room) and whole-house level. If a dataset is listed with several aggregation levels that typically means that the more fine-granular level

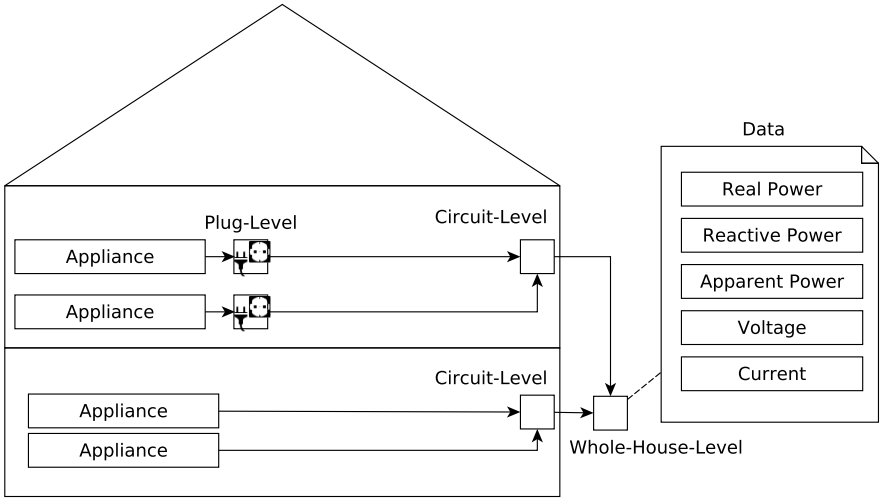


Figure 3.1: Measurement methodologies of the listed datasets.

is only partially available (e.g. for some appliances).

Among the datasets, especially Individual Household Electric Power Consumption (IHEPC), UK Dale, REFIT and Almanac of Minutly Power (AMPd2) datasets captured energy demand over a long time. Hébrail and Bérard (2012) captured the IHEPC dataset with the minutely energy demand of one household for almost four years. In the most recent 2017 release, the UK Dale dataset by Kelly and Knottenbelt (2015), captured the energy demand, including sub-meters and switch status of appliances, of five households for over four years. In 2016, Makonin, Ellert, Bajić, and Popowich (2016) provided, with the AMPd2 dataset, readings for two years of monitoring from 21 power meter and additional climate data. Murray, Stankovic, and Stankovic (2017) recently published REFIT, a two year UK dataset with whole-house level loads as

well as nine separated end-uses measured at a resolution of eight second for 20 residences.

The ECO, Indian Dataset for Ambient Water and Energy (iAWE) and Greend datasets offer high resolution data with 1Hz measurements. Beckel, Kleiminger, Cicchetti, Staake, and Santini (2014) introduced with the ECO dataset power meter readings of six households, updated every second. The set includes manually labelled occupancy data and readings of selected appliances. The Greend dataset provided by Monacchi, Egarter, Elmenreich, D'Alessandro, and Tonello (2014) includes 1Hz measurements of households in Austria and Italy. They monitored the power of 9 households for one year, including nine sensors per house. Batra, Gulati, Singh, and Srivastava (2013) monitored in his dataset iAWE a house with 33 appliances over 73 days.

Smart* and Reference Energy Disaggregation Data (REDD) measured many individual appliances. Barker et al. (2012) observed in the dataset Smart* 25 circuits and 29 appliance monitors over three month. Kolter and Johnson (2011) presented the REDD dataset, which monitored six houses in boston with up to 24 sensors over several weeks. The dataset with the most appliances included is Tracebase. Reinhardt et al. (2012) published with Tracebase, a dataset from Germany with 15 houses and 158 appliances per house in 1Hz resolution. However, Trace-

base does not offer measurements of a coherent duration or aggregated measurements.

This study uses the ECO dataset, provided by Beckel et al. (2014), which is a comprehensive dataset for NIALM and occupancy detection research, offering individual appliance and occupancy readings every second.

The real power (W) values of smart meters are based on the SML-protocol, which captures the mean-cycle-power. The ECO dataset provides, apart from declared exceptions, measurements in Watts with four decimal places for a total length of one year and for six different houses. Due to missing data, the ECO dataset has approximately 120 days (from June 2012 to January 2013) which are simultaneously captured in all households. The experiments throughout the thesis, if not otherwise declared, aggregate the values to one measurement every five minutes, since the lower resolution is more realistic in a smart grid environment. Aside from the whole-house level data with power, current, voltage and phase angle, the dataset contains separate measurements on plug level as well as occupancy data. The occupancy data, which was manually specified by the occupants as presence and absence, was not used in our experiments. The plug level data shows power data for some selected appliances which differ for each household.

Appliance Load Shapes

A load curve consists of the sum of the loads of all active individual appliances and measurement errors. Hence, if a set of appliances and their load curves are given, the total load curve can be reconstructed by choosing the correct state (active/inactive) for each appliance at each time.

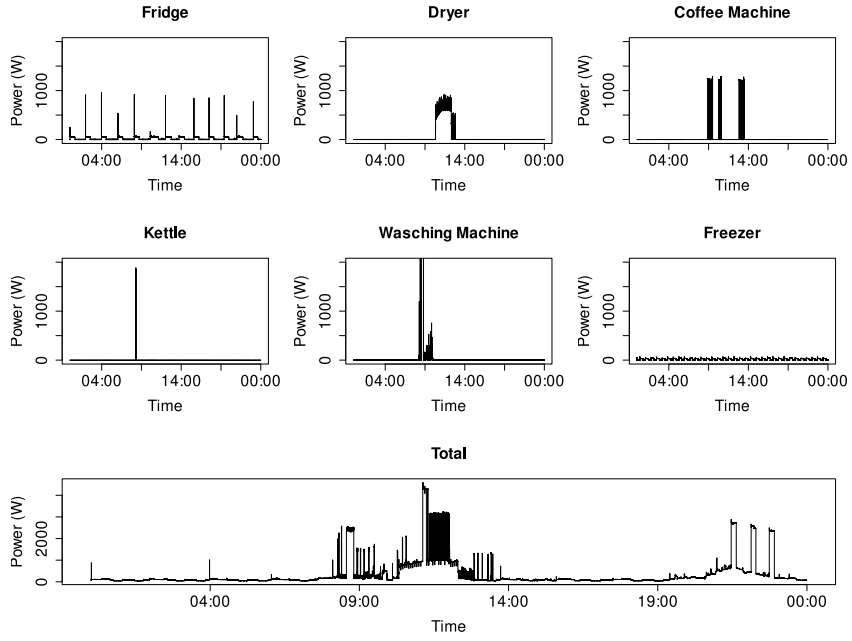


Figure 3.2: Plug level and whole-house level data of one day.

Figure 3.2 shows a snippet of energy demand data, over one day, for the individual appliances and the whole-house level data (bottom) for household 1 of the ECO dataset. Note that, the sum of all individual appliances (e.g. fridge, dryer, coffee machine) is not equal to the measurements on whole-house level. Different appliance types can be identified

by looking at the shapes of the individual appliances load curve. Hart (1992) categorized appliances into 'Type I: On/Off Appliances', 'Type II: State Machines' and 'Type III: Continuously Variable Devices'. The coffee machine and kettle are of type I, while the other appliances are of type II. A type III appliance, such as a personal computer has no fixed load curve and depends on the usage. Other authors, such as Zoha et al. (2012) further extended the types with 'Type IV: Permanent Consumer Devices', which classifies appliances that permanently consume energy at a constant rate (e.g. smoke detector).

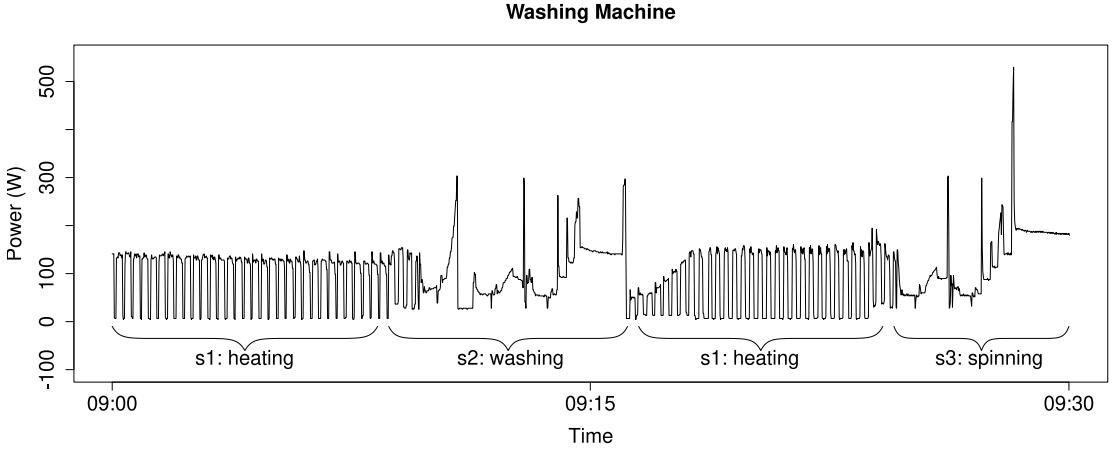


Figure 3.3: Appliance load curve of a (type II) washing machine.

Figure 3.3 shows the structure of a type II appliance extracted from the ECO dataset, with states s1, s2 and s3, in detail. The consumption changes when the appliances switches into another state. The overall load curve of a real world appliance can be modelled as following:

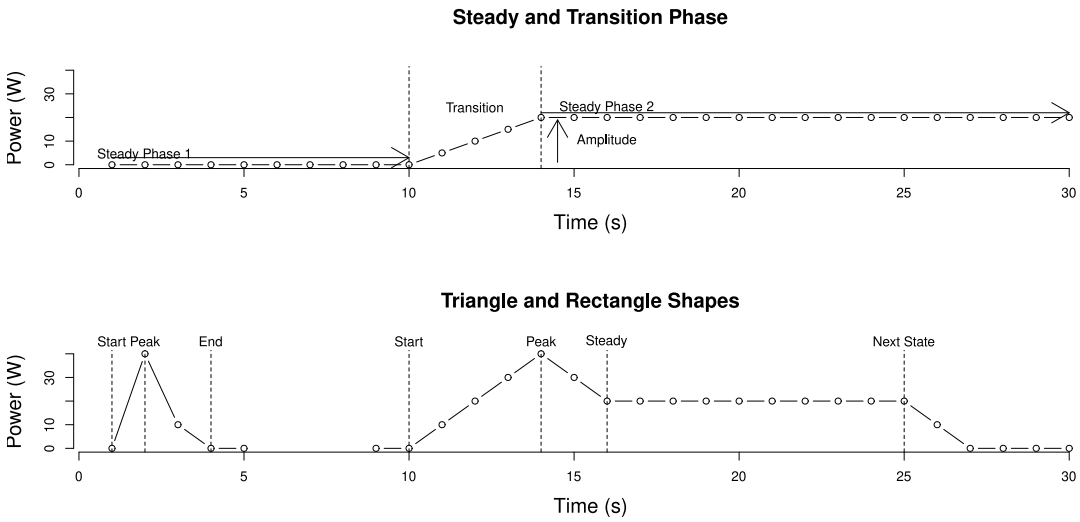


Figure 3.4: Classification of load shapes.

The upper part of figure 3.4 shows the classification by Hart (1992), as previously introduced in Chapter 2, who distinguishes the transition phase and the steady phase of an appliance load curve. The bottom part shows the classification by Z. Wang and Zheng (2011), which enhances the shape by a peak value before the steady phase. Additionally, the classification distinguishes triangular shapes in case of fast switching events and rectangular shapes with a steady phase.

Overview of Modelling Methods

Certain methods require the testing of unorthodox customer profiles or and appliance usage data for different scenarios, which are not always easily available in real world data. For these reasons, the simulation of appliance load profiles with different appliances and occupancy scenarios

Table 3.2: List of energy demand models.

Name	Year	Resolution	Method
Walker et al.	1985	15min	Time-of-use
Train et al.	1985	1h	Probability
Capasso et al.	1994	15min	Time-of-use
Yao et al.	2005	1min	Time-of-use
Strokes	2005	1min	Probability
Paatero et al.	2006	1h	Probability
Widén et al.	2009	5min	Time-of-use
Armstrong et al.	2009	5min	Time-of-use
Richardson et al.	2010	1min	Time-of-use
Arshad et al.	2013	1min	Probability

is very attractive.

Table 3.2 lists ten methods to model energy demand and their corresponding properties in chronological order. The table shows that the research on energy demand models is older than the smart grid, but there is a drift to high resolution models after the very early publications. Modelling methods are typically categorized as top-down or bottom-up methods. Top-down methods often use socio-economic data derive the consumption of a single household from factors such as location, number of persons and income. Bottom-up methods simulate individual appliance load curves which contribute to the total energy demand of a household.

Top-down methods are less interesting for this work, because the shapes of individual appliances are typically not accurately modelled. Such methods are primarily used for rough long-term predictions. High resolution energy demand simulators are almost always bottom-up ap-

proaches, whereas each appliance is typically modelled with electrical characteristics and a start and end time.

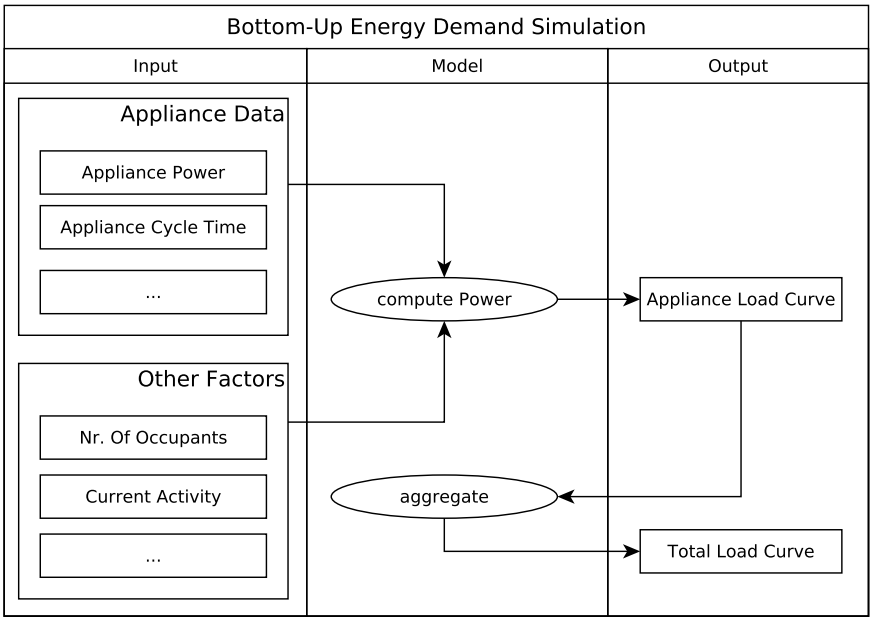


Figure 3.5: Simple bottom-up energy demand simulation approach.

Figure 3.5 illustrates a simplified version of this concept. Input parameters range from the individual consumption of the selected end-uses and their electrical characteristics to the properties of the dwellings, time of the year, weekday and human behaviour. Grandjean et al. (2012) pointed out that bottom-up simulation approaches can be further categorized in time-of-use based or probabilistic models. Time-of-use based models are an extension of probabilistic models, which additionally model the presence or activity of human beings.

Probabilistic models are often rather simple, Paatero and Lund (2006)

used specific appliance consumption data, such as on/off amplitudes as well as the length of run cycles to generate appliance profiles by starting appliances until the profile matches the expected daily energy demand. Stokes (2005) used the mean household demand and dwelling specific data (e.g. number of occupants, number of appliances, ...) to calculate the probability of an appliance start event at a given time. Train, Herriges, and Windle (1985) statistically adjusted generated appliance load curves according to socio-economic factors and Arshad, Ali, and Javed (2013) computed an appliance usage probability, according to different device categories.

While the previous simpler models make use of the appliance start time alone and generally use random number generation to start or stop appliances according to pre-computed probabilities, the state of the art approaches use sophisticated time-of-use models for certain human activities, which are correlated to a set of appliances. These occupancy models are often derived from empirical surveys and are sometimes referred to as 'time of use' based modelling. Walker and Pokoski (1985) were the first to generate a time of use model to start appliances. They calculated an 'Available At Home Probability' and based on that the probability to perform some activity which indirectly affects the use of one or more electrical appliances. Capasso, Grattieri, Lamedica, and

Prudenzi (1994) developed ARGOS an improved version of Walkers time of use model. Widén et al. (2009) generated a time of use information from surveys. Yao and Steemers (2005) used constructed time of use scenarios to start individual appliances with their so called 'Simple Method Load Profile'. Armstrong, Swinton, Ribberink, Beausoleil Morrison, and Millette (2009) created a time of use based model for Canadian households. Richardson, Thomson, Infield, and Clifford (2010) generated Markov-Chains to model the time of use for several appliance categories. The term 'time of use' is coined by works related to energy modelling and simulation, but throughout this work, 'time of use' is referred to as 'human activity' which fits better in our context.

In summary, public datasets are convenient to evaluate against common benchmarks, since the collection of real world data is a tedious task. A quick glance at the energy demand shows that the total load is the sum of each individual appliance load, and hence it is feasible to forecast the total load by predicting the number of appliances switched on at any given moment. The load curve of an individual appliance can be classified into a transition phase and a steady phase, whereas the difference of power is called amplitude. Appliances such as state machines may have different amplitudes and a load curve consisting of several switch operations. The switch operation of each individual device is usually

defined by time of use probability – or simply the human activity.

3.2 A Benchmark for Characteristic Periods

In order to detect anomalies, such as energy theft, it is first necessary to mathematically define the normal consumption behaviour of a household. A solid theoretical framework to understand underlying mechanics of daily energy demand distribution in several dimensions – across several days as well as multiple households – is still missing. This section provides an analytic approach to understand the distribution of power throughout the day and a quantitative metric to measure the similarity of energy demand curves. Energy demand contains complex and variable patterns which legitimately change at regular and indeterminate intervals. Patterns that can be used to recognize households, are not uniformly distributed on the temporal axis of the energy demand. Since energy demand in residential households is profoundly affected by human behaviour and occupancy, which is neither random nor uniformly distributed, characteristic energy demand shapes are often gathered in few periods per day and repeat daily according to the consumers habits. Following this reasoning, one can define a characteristic period of a households daily energy demand as period which does not significantly change over several days, but differs from other households energy de-

mand at this time. The experimental evaluation first analyses the proportion of consistent periods of a household across several days and then, examines the proportion of unique periods of a household to distinguish it from other households.

Since two households with identical average consume could have significantly different daily consumption curves, the here proposed strategy is to capture the uncertainty of a consumers demand. The entropy, also known as uncertainty, can be used to quantify the likelihood of a period to be 'characteristic' or reveal significant differences in the magnitude and timing of energy demand among households.

Although entropy originated in thermodynamics, Shannon (1948) argued with his application of the entropy to information theoretic problems – as a quantitative and qualitative technique for understanding uncertainty – that entropy has a deeper meaning. It is well-known in coding theory that the entropy of a discrete random variable quantifies the average length of the encoding of the random variable. Moreover, Shannon's Entropy measures the average uncertainty, also referred to as information content. The uncertainty is maximized when the outcomes of the random variable are equally likely, which corresponds to a uniform distribution.

Remark 3.2.1. *Formally, let* $L = \{a_1, a_2, ..., a_n\}$ *be a dataset where*

$|L| = n$ is the cardinality of L. Moreover, let $x_1, x_2, ..., x_n$ denote the frequency of each element in L in some sequence X of elements from the set and $m = |X| = \sum_{i=1}^{n} x_i$ implies the number of all observations. Then, the entropy $H(X)$ is defined by

$$H(X) = -\sum_{i=1}^{n} \frac{x_i}{m} \cdot \log_2\left(\frac{x_i}{m}\right) \tag{3.1}$$

Note that, $0 \leq H(X) \leq \log_2(n)$ where $H(X) = 0$ is assumed if only one element of L occurs in X (by convention, $0 \cdot \log_2(0) = 0$) and $H(X) = \log_2(n)$ if all elements of L occur in X with the same frequency. As $\frac{x_i}{m}$ is synonymous for the occurrence probability, the interpretation can be extended to the statistical distribution of the underlying data: an entropy close to $\log_2(n)$ reads as 'random', because the elements appear almost equally often whereas an entropy close to zero reads as 'skewed', because few elements appear more frequent.

In the following, an energy demand curve is split into several time windows and compared with the same window of another curve. When comparing multiple households, the more uniform the energy demand at a certain time distributes across these households, the larger the number of consumer sharing this consumption, and therefore the less likely it is to distinguish consumers from others using this time slot. Alternatively,

the less uniform a period is, the higher the likelihood of a period to be unique among all users.

Table 3.3: Example of energy demand with maximized entropy.

Household	Morning	Noon	Evening	Night
House 1	50-100W	10-15W	10-15W	50-100W
House 2	50-100W	50-100W	10-15W	10-15W
House 3	10-15W	50-100W	50-100W	10-15W
House 4	10-15W	10-15W	50-100W	50-100W

For simplicity, consider a small synthetic example, as outlined in Table 3.3, whose records consist of an identifier (household name) and several attributes in form of load information (interval of energy demand during a period). Hence, each row depicts a day of energy demand for a different household. As the energy demand values appear equally often in each row and column, the table depicts a worst case example: the entropy is maximized for each column and each row and cannot be used to unambiguously link the attributes to a certain identifier. It is not possible to determine any time period as characteristic because none of the time periods in the above table has an unique observation not shared with other households during that period. Hence, neither the rows nor the columns can be used to unambiguously identify the corresponding household. It can be concluded, that the likelihood to unambiguously identify a household is in direct correlation with the amount of power measurements.

The more uniform a households energy demand is spread over the day, the harder to predict the consumers current consumption from historic data. Or in other words, when comparing a period of the energy demand with the same period in historic data, the more uniform the distribution, the less information one can derive about characteristic consumption times.

In summary, the entropy can analyse or model the distribution of energy demand across the temporal axis or across several households. These 'dimensions' can be analysed with the objective of isolating households or atypical characteristic consumption times.

3.3 Experimental Evaluation

The simplest mathematically tractable model of energy demand is a histogram of relative energy demand regarded as a probability distribution, on which one may compute the entropy. However, performing the computation on these relative demand values means to apply a scaling operation which is independent from the other time periods. The resulting entropy for a period with demand (0.1W, 0.09W) and second period (100W, 90W) is therefore equally high.

As an alternative, the author proposes to define a set of intervals on the power and compute the probability to see energy demand in a

Table 3.4: Computing the entropy on power (W) intervals.

Time	0-100 (W)	101-500 (W)	... 1000-∞ (W)	Entropy
1	[Day1, Day2]	[Day3]	... []	0.7
2	[Day1]	[Day2]	... [Day 3]	1.5
3	[]	[]	... [Day1, Day2, Day3]	0.7
4	[]	[]	... []	0.6

certain interval, as depicted in Table 3.4. Here, each column shows a different interval, whereas each row shows a certain time. The entropy is computed for each row: for each interval, the number of days is counted, which contain measurements in this interval. A high entropy means that the intervals are equally distributed throughout the days, and therefore, that none of the intervals is characteristic for this row.

Remark 3.3.1. *As the entropy in the previous plots seems correlated to the Standard Deviation (sd) of all days, the following example briefly highlights the differences of both: to compute the sd of a time period for all days, the sd would increase with the difference between two values, whereas the entropy does not consider the distance between values. Consider two time periods* t_1*=(1W, 1W, 1000W, 1000W) and* t_2*=(1W, 250W, 750W, 1000W). The sd for* t_1 *is higher than the sd for* t_2 *, but (under the assumption that each different value is in a different interval) the entropy for* t_1 *is smaller than the entropy for* t_2*.*

Next, the proposed metric is computed for several days of the same household, in order to measure the consistent periods.

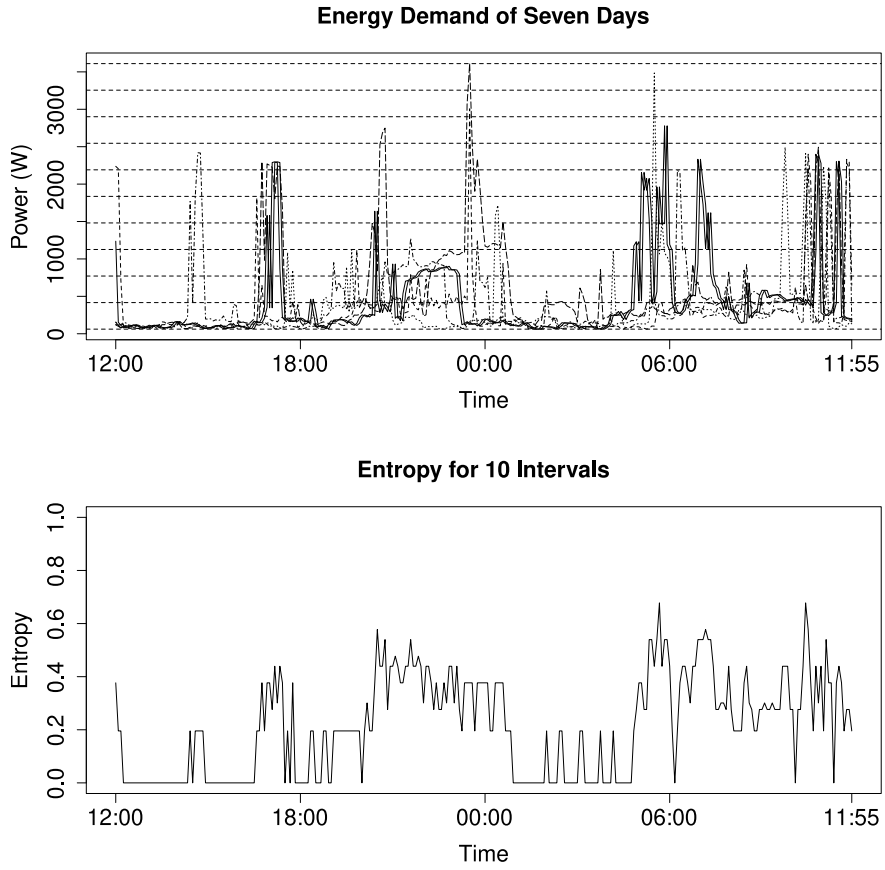


Figure 3.6: Energy demand of seven different days (10 intervals).

Figure 3.6 illustrates the energy demand and corresponding entropy of seven days in a household – the upper plot shows each day visualized by a different line type, whereas the bottom plot shows the corresponding entropy using 10 uniformly distributed intervals. A high entropy means that the values are distributed over several intervals, while a small entropy means that most values are located in a single interval. Here, small entropy values can be seen especially during low activity times, which are more consistent. In contrast to the intuitive assumption that

a household can be identified by peak demands, it is much easier to identify periods without consumption as characteristic (minimized entropy) for a household. The usage of periods without consumption makes sense to characterise load curves: the detection of recurring low consumption periods can be very simply over several days and may be attributed to regular working hours or rest periods, while the identification of a particular appliance, which is used regularly, can be difficult as the consumption may depend on the usage and other hard to predict outside factors.

Note that, the result depends on the number of intervals. With more intervals, there is generally a higher probability that each day has unique intervals, and therefore the entropy is typically higher. The maximum range of the entropy also depends upon the number of defined intervals. The experiments here use a normalized entropy in a range of [0,1]. Interested readers can find another analysis of this parameter in Chapter 6.3.

Figure 3.7 shows the amount of characteristic periods with different intervals, comparing five minute intervals of seven different days. Each box plot in the upper figure shows the result with different amount of intervals (see x-axis). It can be seen here that very small amounts result in a very small average entropy with outliers at peak times. With

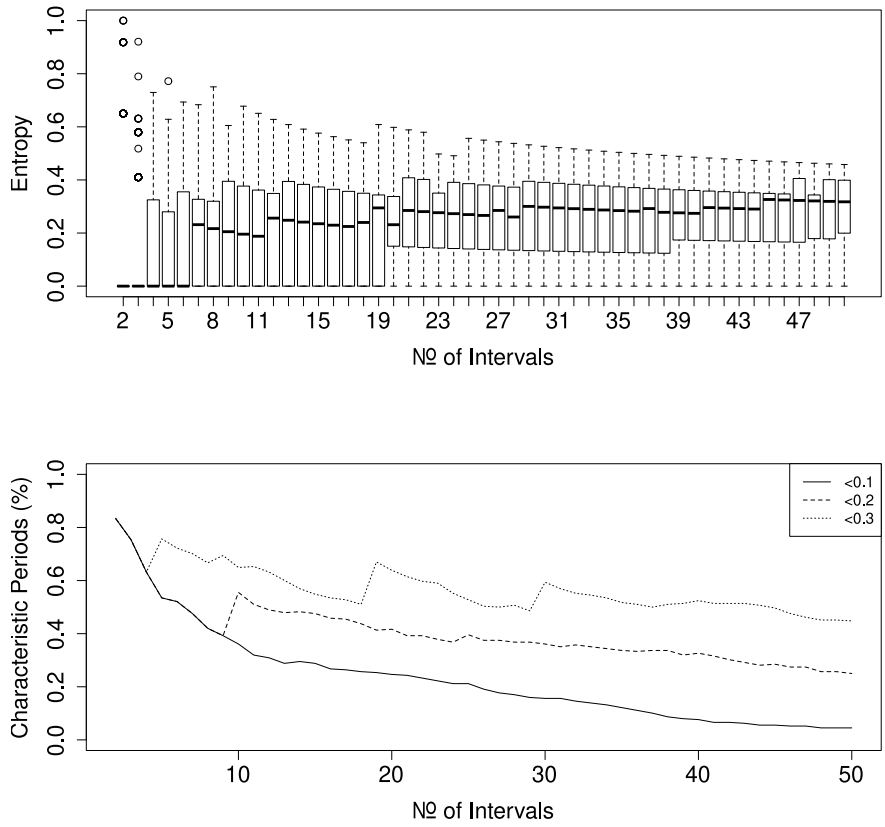


Figure 3.7: Influence of the interval number on the entropy.

increasing intervals, the average entropy also increases. The bottom plot visualizes the amount of five minute periods (%) exceeding the threshold of an entropy of 0.1 (black line), 0.2 (dashed line) and 0.3 (dotted line). The more intervals there are, the less likely it is that all measurements are located in a single interval. The size of the difference does not matter, the entropy will increase whether the intervals are wide-spaced or neighbouring. Hence, with a sufficient number of intervals the entropy of a floating point number may always approach the maximum.

In many cases, two intervals (e.g. high and low) would be sufficient to reliably detect periods without activity. However, in this case the intervals should not be equally distributed, but much smaller for the low activity. For simplicity, the following experiments continue to use a small amount of equally distributed intervals which has the same effect.

The next experiments showcase the entropy for several households, in order to measure the unique periods.

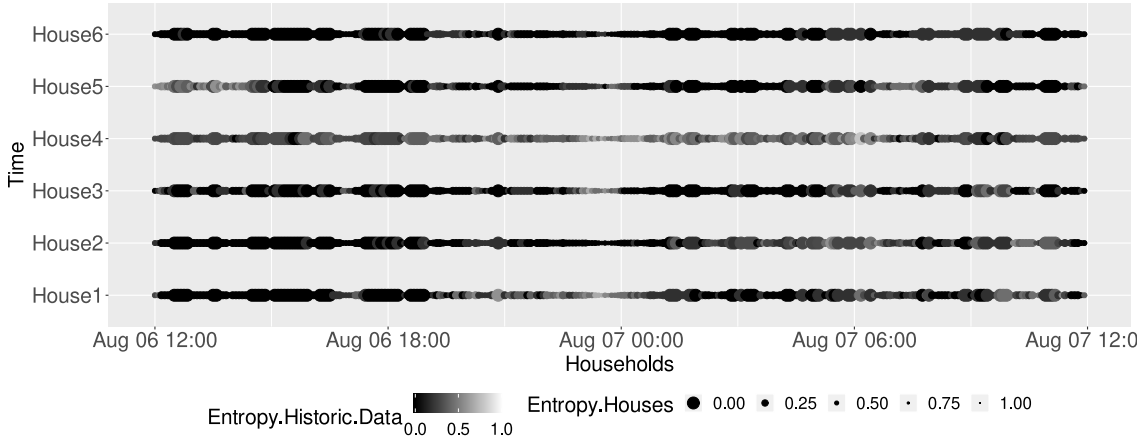


Figure 3.8: Unique and characteristic periods of six households.

Figure 3.8 visualizes the results of both – unique periods and characteristic periods – in a matrix, where each column indicates a time window and each row shows one household. The size of each point shows the entropy between the households and the brightness of each point shows the entropy between days of the same household. The darker the point, the more characteristic is a period, e.g. because several previous days had energy consumption in the same interval. The smaller the point,

the more similar are all households, which means that there is no outlier where one household consumes much more energy and it is difficult to distinguish the households.

In summary, one can compare energy demand with the demand from previous days in order to determine whether this period is random (high entropy) or a characteristic pattern (small entropy) and one can compare the energy demand of several households to find outliers which distinguish the household. Observations of this data suggest that time windows with small energy demand and low activity often display the sleeping habits and regular working hours of occupants, and hence show characteristic patterns of a household.

These findings can help to better structure the data for anomaly detection. Since anomaly detection is, colloquially speaking, the detection of unusual behaviour, it may be argued that anomalies cannot be detected if the demand is not predictable. Hence, it could make sense to either ignore unpredictable time windows or detect whether the predictable time windows occur at the expected time.

3.4 Discussion

This chapter introduced an extensive analysis of energy demand and possible choices of data to detect anomalies such as energy theft. Start-

ing with a selection of popular measurements and available datasets, the chapter introduced common energy demand models which are often based on individual appliances and human activity. Furthermore, an approach to analyse the predictability of energy demand, by comparing previous energy demand or other households and pointed out characteristic time periods, was presented. It is apparent that, especially periods without consumption are often characteristic reappearing patterns of a household. The analysis of two characteristics revealed an answer to RQ2. The first characteristic is the human activity, exposed by time periods with low activity, that can often be identified during the same time spans over several days and can indicate typical consumer behaviour. The second characteristic is the load curve of individual appliances. Each appliance can be classified into a different type such as on/off or state machine and consists at least of a steady and a transition phase. Many authors used these characteristics together to simulate a households energy demand.

Chapter 4

Extracting the Human Activity

4.1 Introduction

The previous chapters illustrated the importance of the human activity to model energy demand and to describe characteristic patterns of a household. Modelling energy demand according to a predefined human activity is straight forward. Next, this study tackles the challenge of extracting human activity from a given energy demand curve. This chapter proposes and evaluates two competing metrics inspired by Shannon's entropy in order to obtain information about occupants' in-home activities and human activity from power (W) readings. The results with an accuracy of over 90%, using the publicly available ECO dataset, indicate that the detection rate with the provided method is significantly better as compared to other well-known statistical methods, stressing the practical relevance of the approach. A focus is especially on entropy-based met-

rics, which are used throughout this thesis, and the advantages and interpretation of using certain input data. This chapter investigates RQ3: 'Is it possible to extract human activity from electric load curves?'. Note that, large parts of this chapter are based on the publication 'Hock, D., Kappes, M., & Ghita, B (2020). Entropy-based metrics for occupancy detection using energy demand. Entropy, 22(7), 731.'.

The central theme here is an effective, entropy-inspired profiling mechanism which consolidates the temporal distribution of energy consumption in order to implement an indicator for human activity. In contrast to most conventional methods this information is revealed without a-priori information, which allows its use without expensive and laborious training of the system. This simplicity is an important factor for the plan to use the human activity as input for anomaly detection, later. Furthermore, the algorithm can also cope with low resolution measurements (minutely).

As human activities that take place without appliances are not reflected in energy demand, this thesis defines the term 'human activity' as human interaction with electrical appliances. Obviously, mispredictions of human activities in this sense are sometimes possible as, e.g., a cleverly randomized use of time-switches cannot be distinguished from a human user doing the same activity. While approaches aiming at

detecting or predicting actual human presence in homes might appear more powerful, the author argues that the differences between these approaches are in fact negligible from a practical point of view when considering the use of such data for anomaly detection and energy forecasts.

Many authors agree that residential energy demand can contribute to the comprehension of household characteristics and human behaviours including the occupancy or human activity. Nguyen and Aiello (2013) give a comprehensive list of approaches to capture human presence, which are mostly based on sensor data. Previous studies of Molina-Markham, Shenoy, Fu, Cecchet, and Irwin (2010) used density-based clustering and supervised learning to identify human presence and other consumer information. Beckel, Sadamori, and Santini (2013) extracted the number of occupants from energy demand. Kleiminger, Beckel, and Santini (2015a), implemented a method based on supervised machine learning algorithms to detect human presence.

Entropy-based approaches have been extensively applied in other areas such as healthcare by Richman and Moorman (2000), biodiversity assessment by Vranken, Baudry, Aubinet, Visser, and Bogaert (2015), or network anomaly detection by A. Wagner and Plattner (2005). To the best of the authors knowledge, an entropy-based metric has not been

proposed to analyse energy demand before. This chapter illustrates and discusses the entropy-based metric as an approach to quantify changes and profile events of domestic daily energy demand and introduce it as a new method to emphasize energy demand variation over time in order to uncover human activity. It seems appropriate to compare two different metrics, which both use raw power values as input data. These metrics can be seen as a filter on energy demand curves resulting in high level information which can be predicted in order to detect anomalies. For the next steps in this thesis, it is especially relevant to forecast this high-level information to detect deviations (anomalies) in the behaviour of a household. For this reason, the interpretation and effects of potential parameters are emphasised.

The remainder, discusses details of the metrics and highlight their advantages and limitations regarding their conclusiveness with respect to residential energy demand. The following section evaluates the method by presenting several practical experiments, obtained by applying entropy to real world smart meter data of the public available ECO dataset. The results show that the proposed metric method can indeed detect human activity. A comparison of the results with other well-known statistical methods, followed by a summary and potential future work avenues, concludes this chapter.

4.2 Entropy as Metric for the Human Activity

Due to appliances with high demand, typical statistical ratios are insufficient to reflect human activity, because the amplitude of energy intensive or energy saving appliances do not necessarily translate into different human activities. Hence, this thesis proposes to use temporal changes of demand, reflected through regularity or randomness, as a benchmark for human activity.

As outlined in Chapter 3.2, one method to summarize the randomness of a variable is Shannon's entropy. The entropy is a convenient way to detect outliers in the regularity of energy demand. The distribution of energy demand can be measured horizontally or vertically. By analysing the distribution of demand over time (x-axis), one may find out that most of the demand appears at a specific time, which is an outlier. By analysing the distribution of measurements (y-axis), one may find out that there are only few measurements with high power, which is also an outlier. The following experiments use the entropy, introduced in Remark 3.2.1, not in a traditional sense but as a metric to mathematically trace outliers in a distribution of measurements. The practical function as a metric, which encodes distributions without losing information on outliers, is the most important aspect for us.

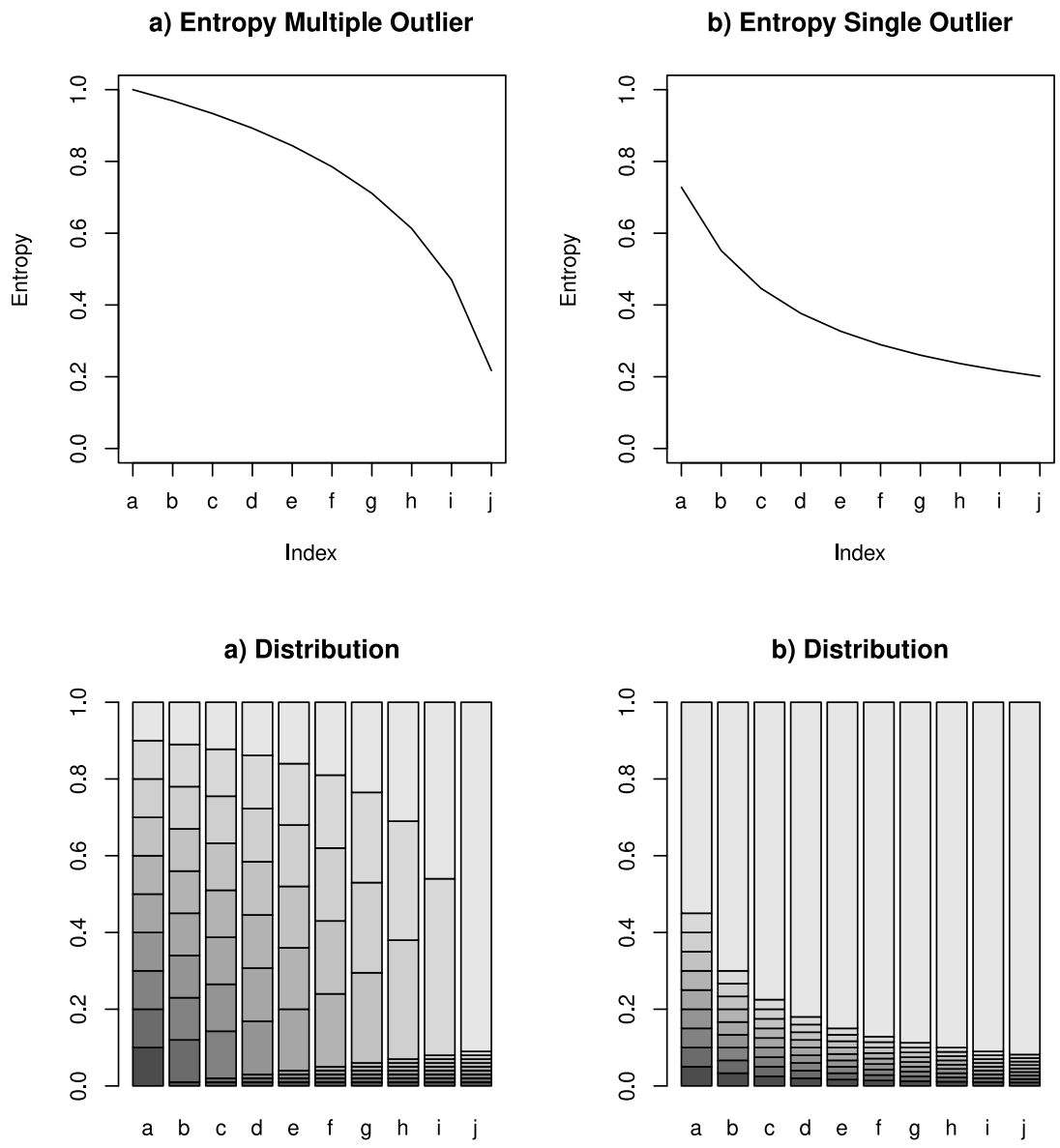


Figure 4.1: Entropies from uniform distribution to skewed distribution.

Figure 4.1 shows the entropy (top) and the corresponding distribution (bottom). The letters on the x-axis correspond to different distributions and the y-axis the corresponding entropy, visualized by the stacked bars

in the bottom plot. On both figures at the left hand side (a), one can see the influence of the outlier number on the entropy. On both figures at the right hand side (b), one can see the influence of the outlier size on the entropy. A distribution such as 'j', with outliers representing the biggest proportion of a vector, is heavily skewed, and hence minimize the entropy.

To conclude: the chapter introduces two competing methods to create the input data vertically or horizontally, which results in two opposite interpretations, different from the information theoretic interpretation. While the first method, *sliding window entropy*, employs the probability to see demand during a certain time, the second method, *interval entropy*, employs the probability to see demand in a certain range.

Sliding Window Entropy

To summarize the sliding window entropy, it utilizes the proportion of energy demand in a time window as input for the entropy to obtain a metric explaining the distribution of energy demand over one day. In contrast to the information theoretic entropy, this metric is maximized if the demand in all time windows is equal and the consumption is distributed uniformly throughout the day and minimized when the consumption is concentrated in a single interval. Hence, this chapter

assumes human activity when the entropy is minimized.

Remark 4.2.1. *Consider a finite time series* $\mathcal{T} = x_1, x_2 \ ... \ x_n, \ x_i \in \mathbb{R}_0^+ \ forall \ 1 \leq i \leq n$, *with* n *elements – representing energy demand. Then, a sliding window* $\mathcal{X}$ *with size* m, *where the sliding step* $(1 \leq \Delta \leq m)$ *defines the number of elements by which the window slides each iteration* t, *can be defined as* $\mathcal{X}_t = x_{t\Delta}, x_{t\Delta+1}, ..., x_{t\Delta+m-1}$, *which results in a total of* $0 \leq t \leq \frac{n-m}{\Delta} + 1$ *windows to cover all elements. The traditional entropy considers the occurrences* x_i *over the data length* n, *which is equivalent to the probability to see a certain element* $\frac{x_i}{n}$. *In energy usage context, the entropy considers the demand probability in a sliding window* $P(\mathcal{X}_t)$ *which is equivalent to the normalised integrated area under the demand line. The approximation of the area is simplified by dividing the demand of a window* $h(\mathcal{X}_t)$, *which is the sum of all measurements in a window, by the total demand* $h(\mathcal{T})$*:*

$$P(\mathcal{X}_t) = \frac{h(\mathcal{X}_t)}{h(\mathcal{T})} \tag{4.1}$$

The entropy is maximized if the demand is distributed uniformly and minimized if the total demand is present only in a single sliding window. The result ranges from $[0, log(\frac{n-m}{\Delta} + 1)]$. *Equation (4.2) illustrates how*

to calculate the entropy using the above probability.

$$H(\mathcal{X}) = -\sum_{t=0}^{\frac{n-m}{\Delta}+1} P(\mathcal{X}_t) \cdot \log(P(\mathcal{X}_t)) \tag{4.2}$$

It is also possible to interpret the probability as the proportion of daily energy used during that sliding window. Note that, the sum of all probabilities can be greater than 1, if the sliding windows overlap ($\Delta < m$), which results in a scaled entropy. Figure 4.2 illustrates the energy demand (grey line) divided into three time windows $\mathcal{X}(1:3)$, $m = 8h$, $\Delta = m$ (dotted lines), with the probability for energy demand delineated in black.

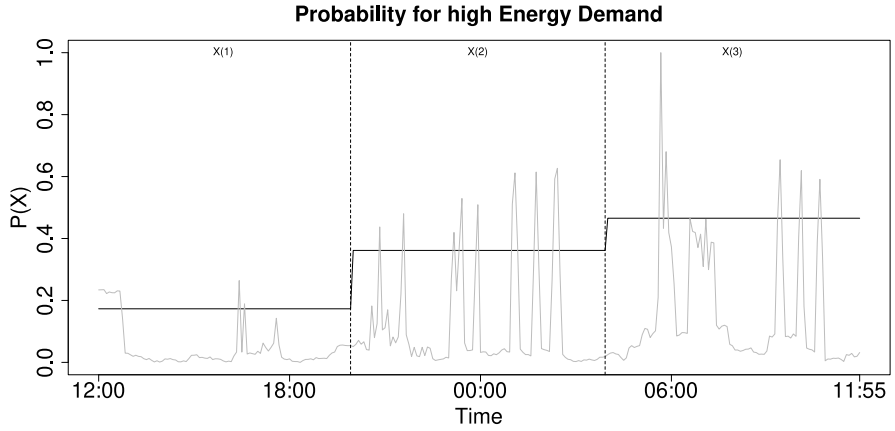


Figure 4.2: Integrated demand for three time windows.

Human activity stands out due to the imbalance and change of energy consumption, which manifests here in form of demand peaks. High demand concentrated in a single time window results in a minimized sliding window entropy. Figure 4.3 visualizes the complete process to

derive the human activity from raw energy demand in 8 hour steps. The figure first presents the probability for high demand (middle) in three different resolutions and then shows the entropy for a time window of three hours using the probability as input data.

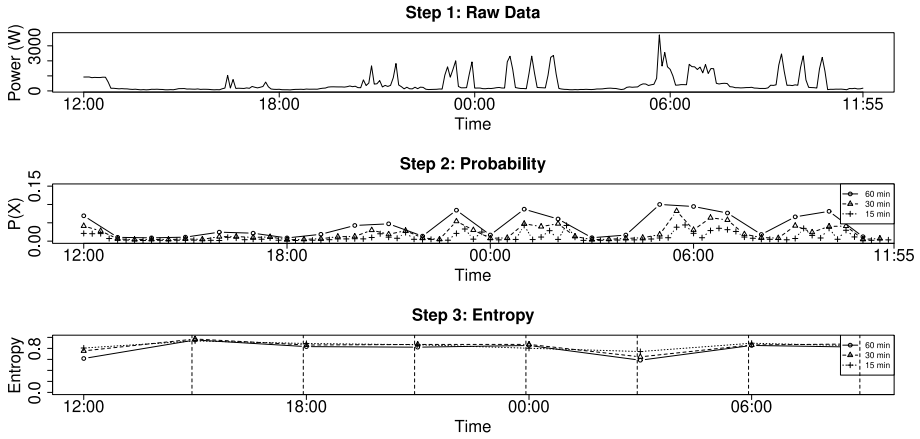


Figure 4.3: Probability to use energy in a certain time window.

The resulting plot highlights the influence of several parameters, namely the number of time windows used as input for the entropy (middle) and the size of the time window for each entropy value (bottom). The resulting entropy is smaller the more input values there are, while the pattern of the curve is unchanged. The entropy results in a filter operation over the energy demand and smaller time windows for the entropy would show more detail. There is a visible drop of the entropy whenever the amplitude significantly changes within a time window. For the decision, whether a time window represents activity or not, a threshold can be utilized. What may look unusual is, that two time windows with

$(0.1W, 1W)$ and $(1W, 100W)$ result in the same entropy.

Interval Entropy

To summarize the interval entropy, the method uses the amount of measurements in each interval to compute a probability resulting in a metric affected by the distribution between intervals. This metric is maximized if the demand values are uniformly distributed over all intervals, which is the direct opposite to the previously introduced sliding window entropy.

Remark 4.2.2. *Consider a finite time series $\mathcal{T}$ (see Remark 4.2.1), with n elements. Then, a finite number m of equal-sized discrete intervals $\mathcal{I}$ covering a total range of $r = max(\mathcal{T}) - min(\mathcal{T})$, is given by $\mathcal{I}_i = [\frac{r}{m} \cdot (i-1) + min(\mathcal{T}), \frac{r}{m} \cdot i + min(\mathcal{T})]$, with $1 \leq i \leq m$.*

The entropy represents the probability that the demand is an element from a certain interval $\mathcal{I}$, where n is the number of elements in $\mathcal{T}$ and m is the number of intervals. The probability of each interval $P(\mathcal{I}_i)$, to contain energy demand, is given in equation (4.3).

$$\begin{aligned} Freq(\mathcal{I}_i) &= \sum_{j=0}^{n} x_j \in \mathcal{I}_i \\ P(\mathcal{I}_i) &= \frac{Freq(\mathcal{I}_i)}{m} \end{aligned} \tag{4.3}$$

The entropy is maximized if the demand distributes an equal number of elements to each interval m and minimized if the demand involves

only one interval. The result ranges between $[0, log(m)]$. *Equation (4.4) shows how to utilize the probability of each interval to compute the entropy.*

$$H(\mathcal{I}) = -\sum_{i=0}^{m} P(\mathcal{I}_i) \cdot \log(P(\mathcal{I}_i)) \tag{4.4}$$

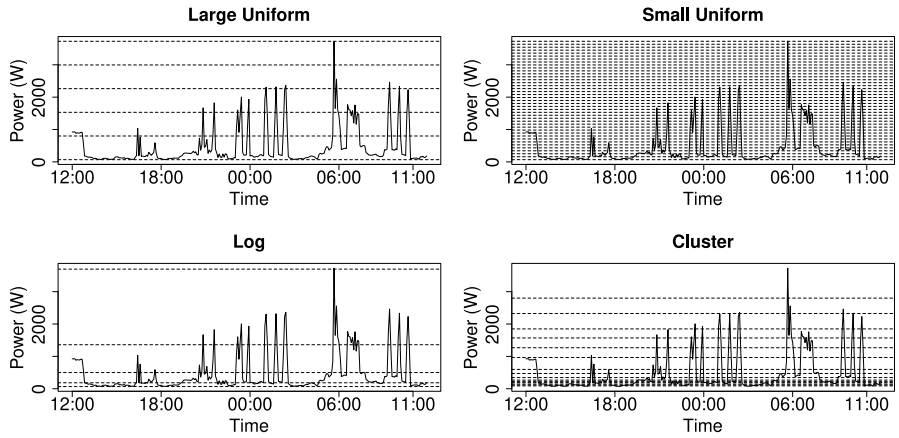


Figure 4.4: Energy demand in different Intervals

Figure 4.4 highlights different methods to generate intervals: the top figures show equally distributed intervals, while the bottom figures show intervals with dynamic size: namely logarithmically distributed and clustered to high variance sections. The dynamic intervals enables the method to also consider small changes of energy demand. However, the author found that large intervals react well on changing variance and reliably show the start and end of each activity phase, while too many small intervals may emphasize changes which do not necessarily

depict human activity. The clustering approach appeared to be difficult to implement in real world scenarios, since the clusters differ for each household and the distribution of cluster centres may show legitimate changes each month.

4.3 Experimental Evaluation

The following experiments aim to detect human activity using real world measurements and compare the introduced method with other well established prediction schemes. The section first introduces a method to create a ground truth by labelling human activity using appliance level data. This ground truth data is later predicted without access to the appliance level data. The author is well aware that the presence of a reliable energy disaggregation algorithm would label the method futile. However, at present energy disaggregation methods are quite complex and, outside the laboratory environment, often not able to reliably detect appliances which are not included in a hand-labelled appliance database.

The experiments use the previously introduced ECO dataset and in order to create the ground truth data, utilizes the ECO appliance level data of one from six household over 30 days (June – July 2012), which includes seven appliances (Fridge, Dryer, Coffee machine, Kettle, Washing machine, PC, Freezer). Events which point to human activity are

labelled by selecting appliances with non-periodic behaviour (4.5c) and exclude appliances in stand-by (4.5a) mode, appliances running permanently and appliances with regular energy pattern (4.5b) (e.g. freezer, fridge, ...).

Consider each of the i appliances in the dataset as finite time series $\mathcal{T} = x_1, x_2 \ ... \ x_n, \ x_i \in \mathbb{R}_0^+ \ forall \ 1 \leq i \leq n$, with n elements. Each appliance $\mathcal{T}_i$ that meets following conditions is selected:

$$C_1 = max(\mathcal{T}_i) > 10 \tag{4.5a}$$

$$C_2 = {}^1Q2(\mathcal{T}_i) - min(\mathcal{T}_i) < max(\mathcal{T}_i) - Q2(\mathcal{T}_i) \tag{4.5b}$$

$$C_3 = {}^2Q3(\mathcal{T}_i) - Q2(\mathcal{T}_i) < max(\mathcal{T}_i) - Q3(\mathcal{T}_i) \tag{4.5c}$$

If an appliance meets these conditions, then all values greater than the arithmetic mean value of the appliance are labelled as human activity.

Activity detection without historical data is not a well researched topic with many approaches available, therefore the introduced method is compared to approaches often used for time series analysis. With the data outlined above, the method is compared to four well-known algorithms, namely a) TsOutlier C. Chen and Liu (1993) , b) Moving Average with an threshold, c) Seasonal Decomposition Kendall and Stu-

[1] Q2: the second quantile is the median.
[2] Q3: the third quartile is the middle value between the median and max.

art (1983) with a threshold on the trend component and the d) standard deviation with a threshold. Each algorithm aims to classify human activity (true or false) for 15 minute time windows, whereas a) classifies windows as 'true' if it contains an additive outlier, b) if it has a demand higher or lower half the standard deviation of the moving average, c) if the trend component contains values higher than the mean, and d) if the standard deviation is higher than the daily main. The two, in this work, proposed algorithms are evaluated as follows, the sliding window entropy is using five minutely measurements to compute a probability and three values to compute the entropy for the 15 minute time window, whereas the interval entropy is using 20 uniformly distributed intervals. These values were found with manual parameter tuning as in the previous section. However, since the exact values may only apply to this data set, the explanation has been narrowed down to the general effect of each individual parameter instead of all combinations. An evaluation of the parameters with regards to the detection rate can be found in the next experiment at the end of the chapter.

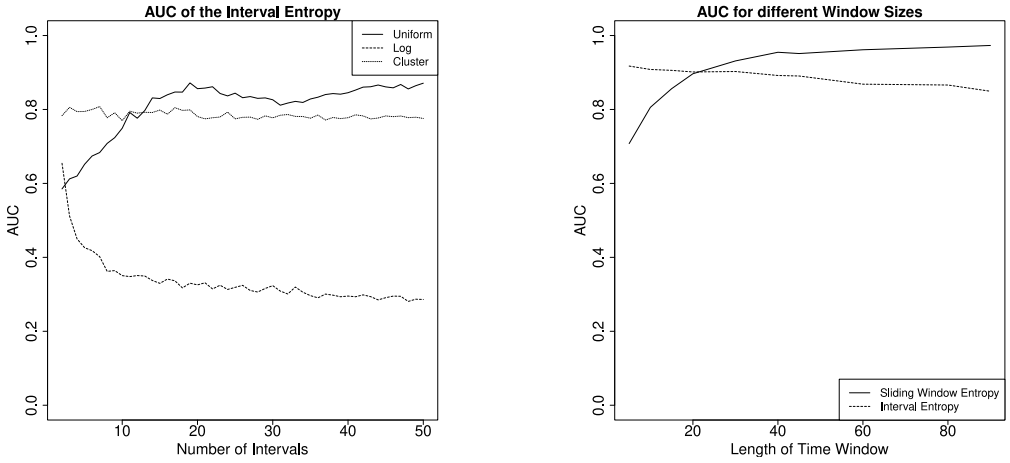
By evaluating the 96 results with the expected ground truth, the Area Under the Curve (AUC) of a Receiver Operating Characteristic (ROC) curve (see table 4.1) is computed. The AUC was introduced in Chapter 2.3 as evaluation metric, where a random method results in a diagonal

Table 4.1: Detection rate of all methods.

	TsOut.	Mov. Avg.	Decomp.	sd	Int. Ent.	Slid. Win. Ent.
2012-06-01	0,57	0,72	0,86	0,79	0,90	0,95
2012-06-03	0,36	0,62	0,80	0,73	0,89	0,95
2012-06-04	0,37	0,94	0,78	1,00	0,95	1,00
2012-06-05	0,60	0,75	0,74	0,83	0,82	0,95
2012-06-06	0,58	0,84	0,88	0,89	0,92	0,98
2012-06-08	0,59	0,75	0,82	0,91	0,94	1,00
2012-06-09	0,61	0,87	0,69	0,89	0,92	0,87
2012-06-10	0,42	0,77	0,79	0,72	0,71	0,91
2012-06-11	0,59	0,83	0,84	0,86	0,88	0,97
2012-06-12	0,35	0,75	0,87	0,83	0,90	0,98
2012-06-13	0,67	0,78	0,82	0,75	0,78	0,90
2012-06-14	0,58	0,81	0,85	0,87	0,89	0,99
2012-06-18	0,53	0,74	0,94	0,80	0,88	0,96
2012-06-19	0,50	0,67	0,91	0,78	0,85	0,97
2012-06-20	0,69	0,68	0,90	0,66	0,69	0,98
2012-06-23	0,51	0,85	0,81	0,89	0,89	0,94
2012-06-25	0,47	0,73	0,82	0,77	0,93	1,00
2012-06-27	0,56	0,75	0,79	0,85	0,91	1,00
2012-06-28	0,45	0,75	0,79	0,85	0,81	0,95
2012-07-07	0,47	0,81	0,87	0,91	0,94	0,99
2012-07-08	0,50	0,90	0,89	0,99	1,00	0,95
2012-07-11	0,59	0,80	0,82	0,90	0,92	1,00
2012-07-12	0,44	0,91	0,91	0,84	0,94	1,00
2012-07-13	0,63	0,72	0,80	0,82	0,83	0,97
2012-07-16	0,58	0,92	0,91	0,87	0,93	1,00
2012-07-17	0,48	0,83	0,90	0,85	0,83	1,00
2012-07-22	0,62	0,82	0,78	0,99	0,98	1,00
2012-07-23	0,63	0,72	0,88	0,78	0,71	0,90
2012-07-24	0,53	0,69	0,92	0,79	0,71	1,00
Total (Avg.)	0,53	0,78	0,84	0,84	0,87	0,97

line and an AUC of 0.5 while a perfect detection results in an AUC of 1. Table 4.1 shows the performance of each algorithm.

Next, the parameters for the interval entropy and sliding window entropy are evaluated. The same data as above is used to compute the AUC for 30 days for different time windows and in case of the interval entropy also for different numbers and types of intervals. A time window of 60 minutes means, that the entropy is computed 24 times per day.



(a) Different interval sizes with window size 15.

(b) Different window sizes with 20 intervals.

Figure 4.5: Effect of parameters on the interval entropy.

Figure 4.5 shows the influence of parameters on the AUC of the interval entropy in detail. The output varies depending on the number of intervals (left plot), where each line shows a different method to distribute the intervals. Here, the uniform distribution showed the best results and is maximized with about 20 intervals and time windows of 15 minutes. Next, the AUC of the interval entropy (20 uniform intervals) and sliding window entropy is compared with different window sizes. The sliding window entropy performs better with larger time windows, while the AUC of the interval entropy is slowly decreasing with greater window sizes.

The parameter tuning obviously depends on the actual household. However, one can clearly see that especially the sliding window entropy does not depend on fine-granular data, as one measurement per time

window is theoretically enough for this method to work.

4.4 Discussion

This chapter presented a non-intrusive approach to detect human activity without a-priori data. With this chapter, RQ3 can be answered in the affirmative: it is possible to extract human activity from energy demand. The first results compared the feasibility of an entropy-inspired metric using different parameters. These results showed a significantly reduced noise, outlier resilience and clearly visible usage patterns. Both methods reflect changes in energy demand with slightly different purpose: the sliding window entropy accurately shows on/off behaviour and peaks of appliances while hiding less interesting absolute values. The interval entropy highlights high variance sections of a time series while masking static behaviour. Both methods appear to behave well with a moderate number of input values. The experiments utilizing the ECO dataset motivated the real world suitability. In some individual experiments, the standard deviation and decomposition method were able to keep up with the entropy-inspired method. However, in total the suggested method archived results which are significantly better than comparable methods without the usage of complex models or intrusive data acquisition. There is some indication that the 'information content'

is profoundly affected by many energy related events and hence, the entropy should be generally suitable for an eclectic range of classification and identification tasks when combined with appropriate parameters.

Chapter 5

Normalized Characteristics of Energy Demand

5.1 Introduction

Chapter 4 proposed the human activity as a characteristic to define the energy demand of a household and outlined the process of modelling an energy demand curve using the activity and appliances. Furthermore, it showcased methods to extract the human activity from energy demand. Here, the thesis investigates alternative characteristics which can be derived from energy demand and discusses the requirements of characteristics to compare households. After an extensive literature research, the chapter introduces three normalized characteristics, designed to indicate the human activity, and to evaluate the influence and effect of energy theft on them. The contribution of this chapter is an analytical approach

to evaluate the effect of energy theft on the proposed metrics. The focus is on the influence of parameters, such as the window size and threshold to compute the feature. This chapter investigates RQ4: 'Is it possible to use different data sources as expected behaviour?'. Note that, parts of this chapter are based on the publication 'Hock, D., Kappes, M., & Ghita, B. (2020). Using multiple data sources to detect manipulated electricity meter by an entropy-inspired metric. Sustainable Energy, Grids and Networks, 21, 100290.'.

The efficient operation of the power grid critically depends on monitoring the participants, which is accomplished by using measurements collected from meters deployed throughout the grid. Since measurements from individual households or from low-voltage networks were not available in traditional power grids, many early studies used information from external sources, such as socio-economic factors and weather data for applications such as forecast, TLC and NIALM.

With the recent wide availability of load data and rapid advancements in statistical methods, akin to machine learning techniques, one can identify a second trend to improve these methods. Here, the authors filter relevant information from load curves by extracting high level information such as occupants, appliances or dwelling size.

Since most information – from external sources or extracted from the

load curve – can also be used for anomaly detection approaches, this section presents a brief overview with the most important types and the related research areas. For a quick overview, Table 5.1 lists ten methods together with the corresponding input data from external sources, and Table 5.2 shows ten characteristics together with the method to extract them from a load curve. The release years on both tables show that the extraction of data, apart from NIALM, is a much newer topic which only came up with the smart grid.

Table 5.1: List of energy characteristics used as input.

Name	**Year**	**Method**	**Input**
Hirst et al.	1977	Regression	Weather, Income
Mihalakakou et al.	2002	ANN	Weather, Solar Radiation, Amplitude
Chicco et al.	2006	SVM, Ant Colony Clustering	Amplitude
Ruzzelli et al.	2010	ANN	Q/P[3]
Saitoh et al.	2010	Clustering	Peak and Mean Ratios
Kim et al.	2011	HMM[4]	Amplitude, On/Off Duration
Zeifma et al.	2012	Bayes	Change-points
Xiao et al.	2014	K-Means	Wavelets
Wojcik et al.	2019	Random Forest	Current, Voltage
Wang et al.	2019	ANN	Hour, Week, Power

One of the early studies on energy forecast is written by Hirst et al. (1977) used demographic, economic, and technological factors in order to model residential energy usage with regression analysis. Mihalakakou, Santamouris, and Tsangrassoulis (2002) implemented an energy model of a Greek house using an ANN with air temperature and solar radiation as input data. Parti and Parti (1980) initiated the research on a method

[3]Real and Reactive Power
[4]Hidden Markov Model (HMM)

called CDA, which attempts to use regression on individual appliances based on the indication of appliance ownership to estimate monthly residential energy use. Wójcik, Łukaszewski, Kowalik, and Winiecki (2019) evaluated different methods such as decision trees and random forest on a number of features. Y. Wang et al. (2019) used a long short-term memory ANN with current hour, week and consumption to predict the nationwide consumption.

Rather than using household characteristics and socio-economic factors as input parameter for various tasks, the observed load profiles can instead be used to indicate those characteristics. Raw load curves are difficult to compare because the aggregated and overlaying patterns of several components can drastically change and pollute the appearance of a load curve. Characteristics extracted from energy demand aim to filter any unnecessary information. As long as the steady and transition phase of appliances are visible (see Chapter 3.1 Figure 3.4), deriving features such as amplitudes, peaks and on/off durations is an easy task. Recently, the wide availability of load data and rapid advancements in machine learning techniques, resulted in many concepts and theoretical frameworks to extract household characteristics. Individual appliances, presence of users or the size of the household can be extracted from the aggregated load data as provided by a smart meter. The resulting

characteristics go far beyond the traditional classification of industrial and residential consumer. Depending on the methodology used, they include information on dwellings, occupants, appliances, historical energy consumption, climate and macroeconomic indicators such as population, gross domestic product, unemployment, and energy price

Energy providers can use these information to better understand consumption behaviour in order to improve the prediction on certain consumer profiles, suggest optimal energy pricing for particular groups or monitor individual appliances. Furthermore, the information obtained by those methods can help to understand the complex consumption patterns of particular consumer groups, and hence may be useful inputs for anomaly detection.

Table 5.2: List of extracted household characteristics.

Name	**Year**	**Method**	**Information**
Yohanis et al.	2008	Regression	Occupants, Dwelling Size
Duckman et al.	2008	Statistical Analysis	Income, Employment
Price et al.	2010	Regression	Income, Employment
Molina et al.	2010	Density-based Clustering	Household Activities
Kolter et al.	2011	Regression	Building Properties
Beckel et al.	2013	KNN[1], SVM, LDA[2]	Children, Bedrooms, Dwelling Size
Mcloughlin	2013	Time Series Analysis, Clustering	Employment, Children
Carroll et al.	2014	Time Series Analysis	Employment, Children
Kleiminger et al.	2015	PCA, SVM, KNN[1], HMM	Human Presence
Newing et al.	2016	Time Series Analysis	Employment, Children, Income

Table 5.2 shows common characteristics which can be extracted from load curves in contrast to the third party information shown in the pre-

[1]K-Nearest Neighbours (KNN)
[2]Linear Discriminant Analysis (LDA)

vious table. The characteristics are often intended for TLC (e.g. to create consumer profiles with dynamic prices) or to monitor safety and security aspects of the smart grid. A study conducted by Yohanis, Mondol, Wright, and Norton (2008) analysed the influence of the number of occupants and the size of dwellings on load curves. Druckman and Jackson (2008), as well as Price (2010) were able to extract the income and employment status of householders. Molina-Markham et al. (2010) used density-based clustering and supervised learning to identify private information about consumers. Kolter and Ferreira Jr. (2011) analysed the relation between energy demand and building properties such as the number of rooms and the building value. McLoughlin (2013) correlated load curves to the employment status and presence of children. Beckel et al. (2013) extracted the number of occupants. Kleiminger, Beckel, and Santini (2015b) used multivariate methods and supervised learning to detect human presence. Carroll, Lyons, and Denny (2014) and Newing, Anderson, Bahaj, and James (2016) associated energy consumption patterns with particular dwellings, income and number of children.

The extraction of such information means e.g. the mapping of the overall activity or activity at a specific time, found in an energy demand curve, to properties such as the employment status, number of occupants or having children. This is typically done by machine learning algorithms

tailored to this property. However, for anomaly detection it may be unnecessary to distinguish a vacation, night shift or unemployment as long as a change is detected.

5.2 Normalized Activity Metrics

Extracting information, such as individual appliances, socio-economic data and personal behaviour, often requires complex, time intensive approaches and databases with expert knowledge. The regular activities of a household, given by e.g. the number of residents, work hours, and sleeping habits are easy to find and still more consistent than raw data. Therefore, this thesis proposes to numerically characterizes the 'activity' of a time period in a households energy demand. Activity can be seen as an occurrence of a state change for one or many appliances, quantified by the number of visible operations, or simply as consumption, quantified by the power level. As one neither needs to distinguish devices nor find the cause of any activity in order to perform anomaly detection, a consistent behaviour of the features is sufficient for the purpose of anomaly detection. This chapter implements three features according to the above criteria, which all estimate the amount of state changes or consumption during a predefined time window to summarize the activity: the number of high amplitude points (f_1), the number of amplitude

changes (f_2) and the number of similar amplitudes in a row (f_3). All three features are computed by applying a binary classification on each measurement and adding up the number of measurements with positive class. A lower activity results in a higher score, and hence an outlier in case of a manipulated electricity meter. By counting the number of measurements in a time window one always receives results with a fixed range, which are easy to compare and to normalize.

Remark 5.2.1. *Formally, consider a finite time series $\mathcal{T}$ (see Remark 4.2.1), with n elements representing energy demand. Then, each feature is a conditional sum over $\mathcal{T}$, formally a function $f : \mathcal{T} \mapsto \mathbb{N}_0$ with a range $[0, n]$, here represented as Iverson bracket with ε defined as threshold in Watt:*

$$f_1(\mathcal{T}) = \sum_{i=1}^{n} [x_i > \varepsilon] \tag{5.1a}$$

$$f_2(\mathcal{T}) = \sum_{i=1}^{n} [|x_i - x_{i+1}| > \varepsilon] \tag{5.1b}$$

$$f_3(\mathcal{T}) = \sum_{i=1}^{n} [l(x_i, \varepsilon) > \delta] \tag{5.1c}$$

The function $l(x_i, \varepsilon)$ returns the amount of neighbouring measurements in a row that are equal to x_i, if rounded by ε, whereas δ is a

threshold for the number of neighbours. f_1 *and* f_2 *are defined as less than to get a big number in case of energy theft. However, this is only more intuitive and also works the opposite way.*

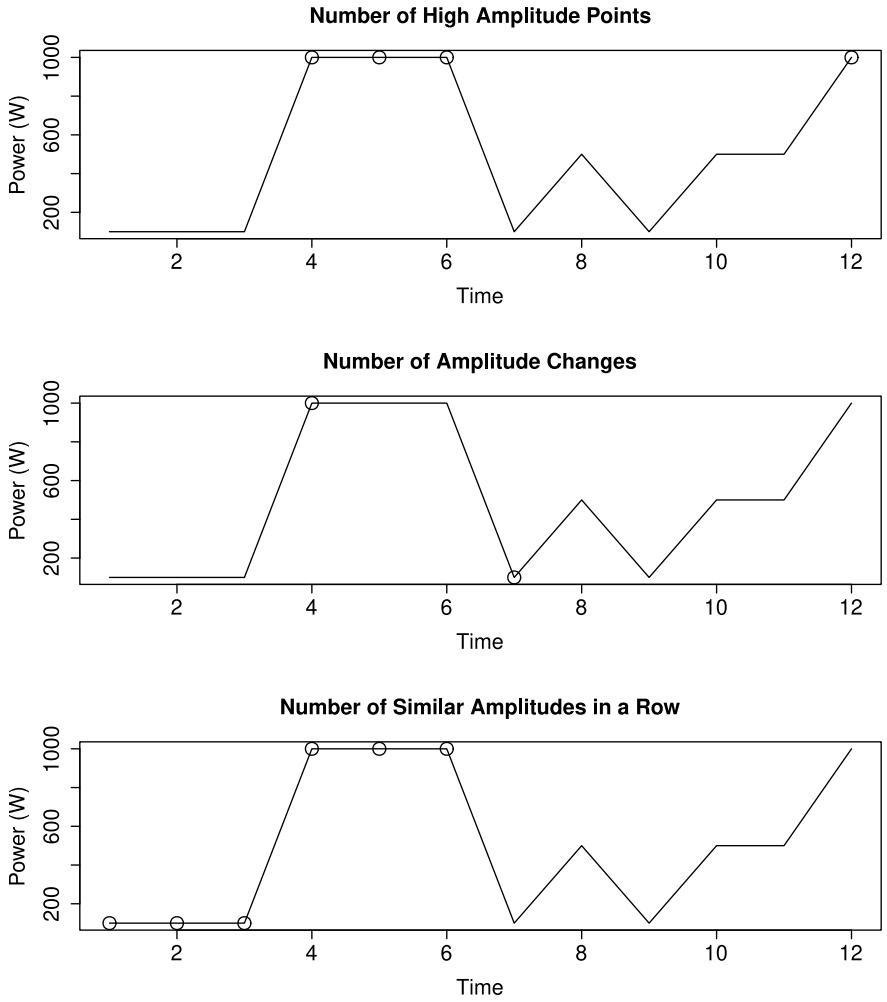


Figure 5.1: Illustration of f_1(top), f_2(middle), f_3(bottom).

Figure 5.1 illustrates the proposed methods to extract features. The line shows an artificial energy demand (y-axis) over time (x-axis), whereas the points illustrate measurements in one of two classes for 'High Am-

plitude' (top), 'Amplitude Change' (middle) or 'Similar in a Row' (bottom).

5.3 Experimental Evaluation

In order to dynamically adjust and optimize the parameters, this chapter introduces some relevant statistical indicators from a holistic data analytic viewpoint, without considering the mechanics of the power grid, which are well-known to operators. It is possible to mimic these manual tuning rules for automation, but the efficient automatized tuning, which is a topic in the realm of optimization algorithms, is not within the scope of this work.

The following section evaluates the significance of the features over raw data and especially considers the influence of the parameters to generate outliers in case of electricity theft. The section aims to provide an in-depth analysis on the effect of threshold ε, the window size, the number of dimensions and the length of training data on the result. The threshold ε can be used to define which deviations from the normal model of a load curve are still within expectations and must be found for each individual electricity meter.

Note that, the following experiments always show the ECO data of household 1 (June 2012 to January 2013), while experiments with sev-

eral houses utilize the data of all six households. In case of missing data the corresponding day is removed from all households, which resulted in approximately 120 days which are simultaneously available for all households. For the experimental setup, the real world measurements from the ECO dataset are utilized to artificially construct cases of tampered with data. In order to motivate a realistic scenario, the data was manipulated according to the traditional (physical) tampering methods introduced in Chapter 2.2. Namely, aiming to bypass energy consumption and slow down the measurement or stop the measured energy consumption altogether. Such patterns are simple, but realistic scenarios. The chapter assumes that the adversary cannot gain unlimited digital access to the smart meter, and hence stealthy energy theft attempts, such as mimicry attacks, which attempt to bypass anomaly detection, are not considered.

Remark 5.3.1. *The **type 1** electricity theft is a time series $\mathcal{T}'_1$ generated by the original data, but with an arbitrary region $\mathcal{F}$, with a length of at least the time window for a feature, replaced by 0 Watt and represents a case where the smart meter is cut off for a certain time. Formally, $\mathcal{T}'_1 = \mathcal{T}[\mathcal{F}] \cdot 0$. Whereas **type 2** is a time series $\mathcal{T}'_2$ generated by the original data divided by a constant $\mathcal{C}$ and represents a case where the smart meter is continuously manipulated to lower the demand. Formally, $\mathcal{T}'_2 = \frac{\mathcal{T}[\mathcal{F}]}{\mathcal{C}}$*

The type 1 falsified data may sound statistically trivial to detect, but regions without power also occur in legitimate load curves (e.g. sleeping hours or working hours), and hence only change the regular load patterns to a decreased activity.

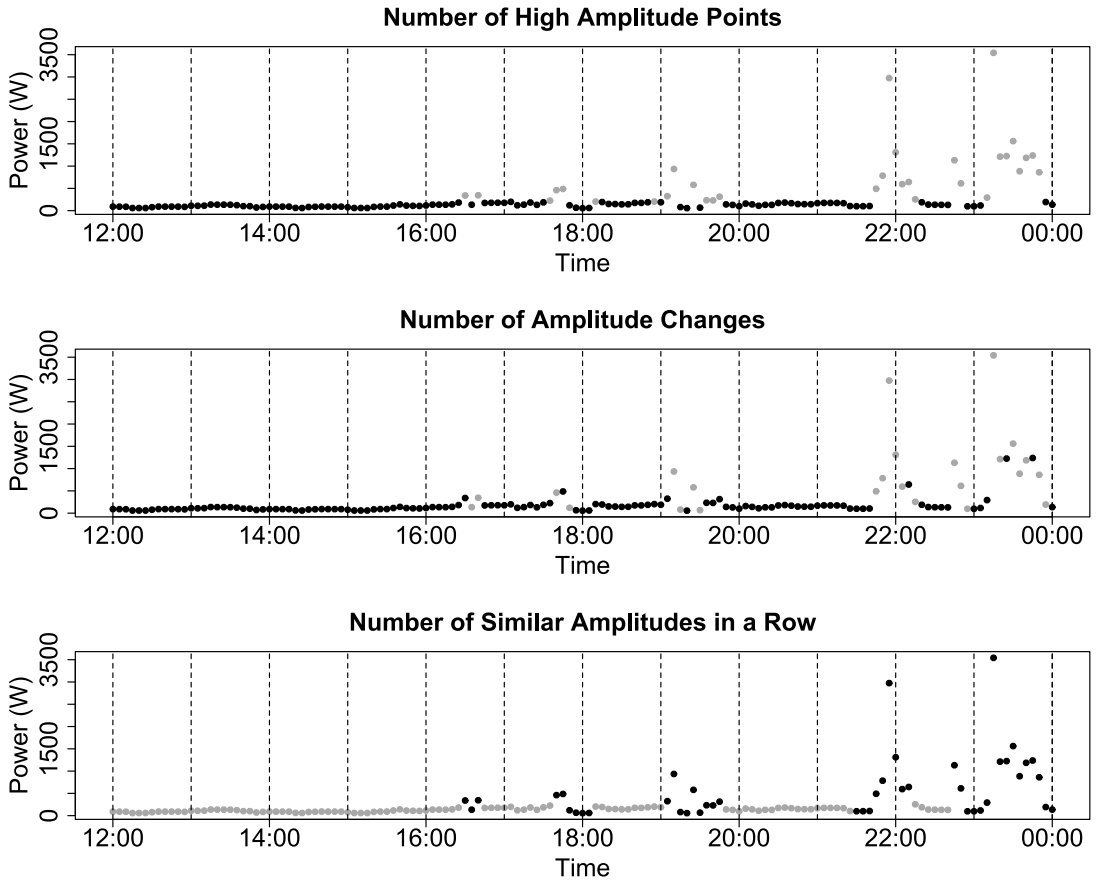


Figure 5.2: Features with threshold $\varepsilon = 200W$.

Figure 5.2 illustrates the features according to the previous section. The plots present the classification of measurements according to the condition of equation 5.1a, 5.1b, 5.1c. Each of the three plots show the power on the y-axis, which is used as input for the proposed methods to

extract features. The grey and black color of each dot illustrates whether the measurement exceeds the threshold of the corresponding method. By counting the number of grey dots, the feature can be computed. The vertical dotted lines visualize the length n of a period which is used to compute the sum over all grey dots.

The quality of ε is assessed in several steps: the first step is to ensure that the features can distinguish anomalous and normal data and can describe the relationship between two data sources. Here, this property of the features is measured using the *correlation*. Furthermore, as each feature is basically a classification of individual measurements, the experiments clarify whether the measurements are evenly distributed over both classes, because a single-sided classification does not contain any information. On top of that, this section illustrates whether measurements of different normal load curves are equally classified or random, depending on ε. In the following, these two properties are called *regularity* and *certainty*.

The first objective is to see the influence of ε on the correlation, which shows the amount of variation that cannot be explained when the features of two load curves are compared. For this experiment, the energy demand of a single day is sliced to equally sized time windows (see Figure 5.2) in order to compute the features, which results in a

vector of features $\vec{v}$ representing energy demand.

Remark 5.3.2. *A correlation of 1 indicates that time series $\vec{v}_1$ reacts at any time exactly like time series $\vec{v}_2$, while a correlation close to 0 means that the two curves are not related. A negative correlation shows that $\vec{v}_1$ and $\vec{v}_2$ is horizontally or vertically reversed. A negative correlation is not interesting for these experiments, because that would roughly mean that $\vec{v}_1$ shows energy production while $\vec{v}_2$ shows energy consumption or that $\vec{v}_1$ shows high consume in the morning while $\vec{v}_2$ shows high consumes in the evening. For this reason, the scale of the correlation (in Figure 5.3) is changed to $[0, 1]$, whereas 0 means different and 1 similar as follows: $\hat{corr} = \frac{(corr(\vec{v})+1)}{2}$.*

The next question to be asked is whether the features of a normal load curve show higher correlation to the features of a normal load curve than to a falsified load curve. This is evaluated by the correlation to falsified demand with different load pattern (type 1: region replaced by 0 Watt) and decreased energy (type 2: region divided by 5).

Figure 5.3 presents the normalized correlation (y-axis) with different ε (x-axis) for each feature (avg. 30 days). The black line plots $\hat{corr}$ of normal data, and should be maximized, whereas the dotted lines each represent the comparison to tampered data, and should be minimized. The plot 'Similar in a Row' is not expected to show a minimized cor-

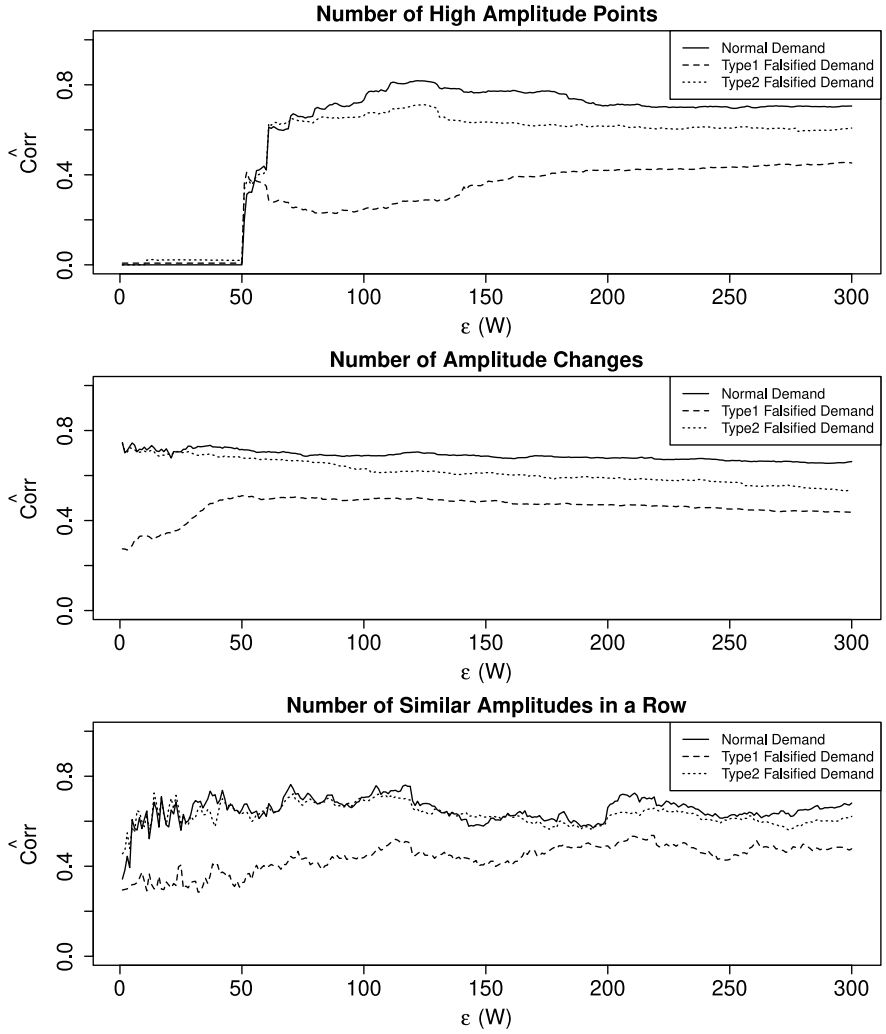


Figure 5.3: Correlation of normal and anomalous load curves with ε.

relation for type 2, because the pattern, and hence measurements in a row, is unchanged through the tampering method. In 'High Amplitude Points', it can be seen that $\varepsilon < 50W$ results in a $\hat{corr}$ of 0 while both other features work with a small ε. This is of course coined to this specific household, which consumes (due to the standby power of many appliances) at least 50 Watt even without human activity. Next, the

computation of *regularity* and *certainty* is introduced.

Remark 5.3.3. *Naturally,* $R = \frac{f_1(\mathcal{T})}{n}$ *shows the ratio between* $\{0, 1\}$ *classifications. Let's define regularity as ratio approaching 0.5, which can be normalized to range* $[0, 1]$*, where 0 is one-sided and 1 regular with* $Regularity = 1 - \frac{|R-0.5|}{0.5}$*. The certainty is used to ensure that the classification is not random. The equation for the feature is used without the sum, which results in a binary time series* $\mathcal{T}' = b_1, b_2 \ ... \ b_n, \ b_i \in \mathbb{B} \ forall \ 1 \leq i \leq n$ *representing one day of energy demand. Let's define a matrix* M*, where each of the* m *columns is an instance of* T' *showing a different day, then a row is 'random', if the* $\{0, 1\}$*-ratio of the* m *instances is 0.5. This chapter defines certainty as the opposite of random,* $Certainty = \frac{|RowRatio-0.5|}{0.5}$ *with a range* $[0, 1]$*, where 0 means random and 1 means that* b_i *is equal for all columns.*

Figure 5.4 (mean of 30 days) shows the relationship of *regularity* and *certainty* (y-axis) for each feature. The x-axis shows different thresholds ε in Watt, which are used to adjust the conditional expression. As only normal data is used in this experiment it is expected that a good threshold would maximize the certainty. A certainty of 1 would mean that the patterns of all 30 days are exactly the same, which is good for anomaly detection because a different load pattern would stand out. The regularity is only used to identify errors, e.g. thresholds with a one-

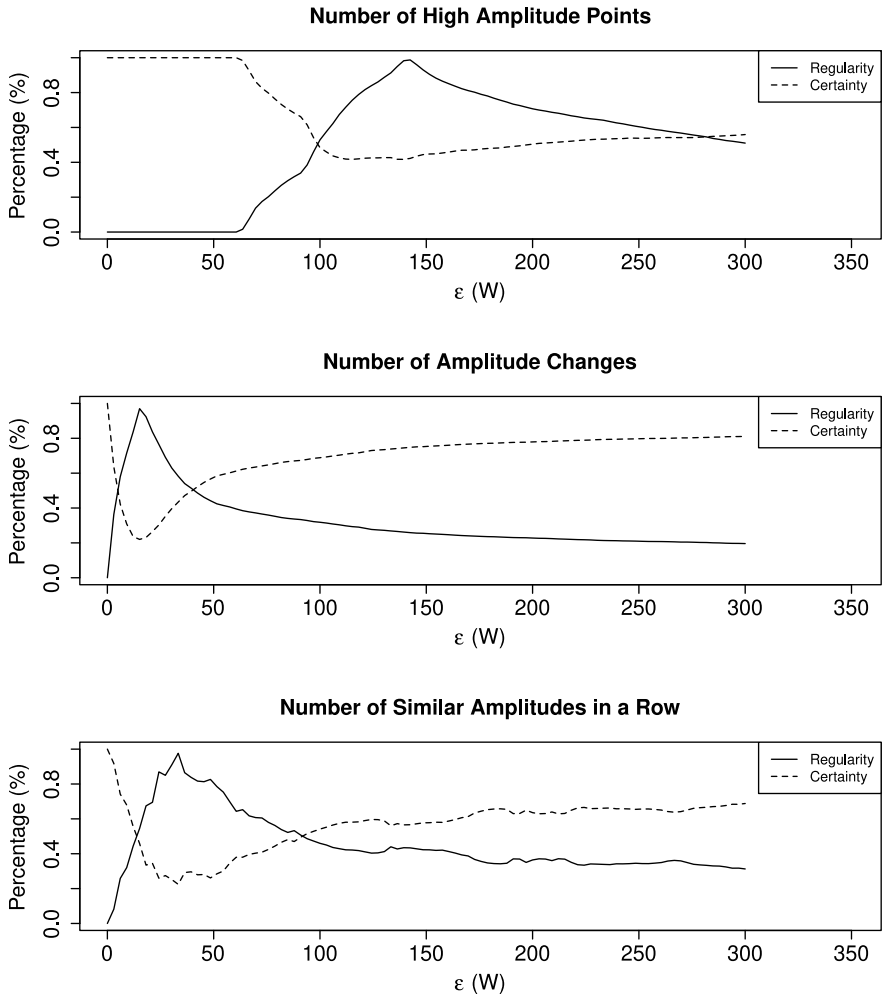


Figure 5.4: Regularity and certainty with regards to threshold ε.

sided classification which can not distinguish between different days at all and for this reason also results in a high certainty. For the anomaly detection itself a low regularity would be perfectly fine. According to this reasoning, one can see that the certainty starts with a high value, but as the regularity at this point is zero, it only means that the feature is unable to distinguish different load curves. Next, a negative peak for

the certainty (approximately for the same ε that maximizes the regularity) can be seen, which should be avoided, as zero certainty means the classification of all 30 days was random. The two intersections of regularity and certainty are a good compromise and show where the explanatory power of the feature is maximized. The result also confirms the estimation, of the previous figure showing the correlation of normal and anomalous load curves, which showed that 'Number of High Amplitudes' needs a high ε (for this household) while both other features work with low thresholds. New to this figure is that it can now be seen that the feature 'Number of Amplitude Changes' results in an overall higher certainty, which means that the patterns of this feature are more stable, and therefore easier to forecast than the other features.

Altogether, it can be concluded that ε can be adjusted, so that the features react well to specific tampering methods or, if the tampering method is unknown, to the energy demand of a certain data source.

5.4 Discussion

This chapter presented an overview of state of the art characteristics of energy demand together with a set of applications. In order to investigate RQ4, the chapter proposed some characteristics that are especially suitable to compare multiple households and showed an experimental

evaluation of these. The prerequisite to utilize multiple dimensions is an outlier producing feature comparable with other data sources. The chapter showed three such features and a systematic approach to fine-tune and adjust them, which significantly affects the detection rate. The chapter demonstrated that high level information, such as the 'activity', can be utilized to normalize data to a fixed range and fine-tune the feature to the specific 'normal' characteristic of a data source. For the fine-tuning, we illustrated how to find an optimal threshold for each feature, to distinguish normal and falsified data. The approach of extensive parameter tuning, to adapt the feature to the specifics of a data source and a certain malicious activity, may be seen as limitation. However, the examination of the statistical influence of parameters was prioritized over automation, because anomaly detection can only work with a solid understanding of the underlying data.

Chapter 6

Anomaly Detection with Multiple Dimensions

Anomaly Detection Systems can discover even unknown and new attacks by analysing statistical deviations from a defined normal behaviour, but what sounds simple is complicated in practice: due to the high sd and inconsistency of load curves, attributed to the on/off behaviour of users and the workings of underlying appliances, separating malicious outliers from legitimate patterns is a challenging task. This chapter proposes and evaluates the benefits of using similar data to detect anomalies. Subsequently, the concept of using different data sources is introduced and an anomaly detection scheme which utilizes an entropy-inspired metric to preserve outliers in multiple data sources to detect electricity theft is evaluated. An advantage of the metric is the robustness against multiple

manipulated data sources, which is a concrete improvement to alternative outlier preserving concepts to aggregate multiple data sources. The detection rates better than 90% demonstrate the effectiveness of using several data sources simultaneously, that, when used individually, provide little value in anomaly detection. The method can complement and enhance existing monitoring systems which usually only analyse a single time series. Furthermore, the proposed method shows that different households can be used as comparable data sources, without clustering the households according to their similarity first. This chapter aims to investigate RQ5: 'Do multiple data sources improve the detection rate?'. Note that, the first part of this chapter is based on 'Hock, D., Kappes, M., & Ghita, B. (2020). Using multiple data sources to detect manipulated electricity meter by an entropy-inspired metric. Sustainable Energy, Grids and Networks, 21, 100290.'.

6.1 Introduction

The following section introduces an anomaly detection scheme, which inspects the power measurements of a smart meter to detect electricity theft. In contrast to traditional anomaly detection approaches, which observe a single source over time to detect tampering, the here introduced approach consults the energy demand of several data sources.

Different data sources are from here on referred to as *dimensions*, because the regular measurements of a smart meter are mathematically considered a finite time series $\mathcal{T}$ (see Remark 4.2.1), with n elements. As such, the time synchronous results of m smart meters are a $n \times m$ matrix $\mathcal{M}$. A dimension, in this works definition, does not necessarily correspond to a physical dimension, but anything that can construct a matrix of comparable data (e.g. it would be possible to add a dimension for: all single-family households, all households of a region, or all Saturdays).

By using similar data, such as historic data or spatially close and structurally identical components that naturally adapt due to the similar conditions, one can mitigate concept drift and unveil otherwise hidden outliers. While it sounds natural that more data is better in order to detect anomalous events, the heterogeneous power data comes together with some challenges which can be addressed by properly sorting and grouping similar data sources.

Remark 6.1.1. *As an example to motivate the structuring of data, consider an anomaly detection model, as introduced in Chapter 2.3, which can assign a certain probability p (or alternatively a numerical score) to each event indicating its abnormality, where p is generated by a function which compares one or more measurements of the current energy*

demand to reference values from historical data or other households. The binary result $r \in \{0, 1\}$ *(where 0 denotes normal and 1 abnormal) depends on a threshold* ε*.*

$$r = \begin{cases} 1 & \text{if } p \text{ is} > \varepsilon \\ 0 & \text{if } p \text{ is} \leq \varepsilon \end{cases} \tag{6.1}$$

If an event was assigned a probability p *above the threshold* ε*, but was actually legitimate traffic, it is call the result a FP. If an event was assigned a* p *below* ε*, but was actually unwanted traffic, it is call the result a FN.*

Figure 6.1 presents a plot with N=100 measurements. The black dots represent normal cases, whereas the grey triangles show malicious cases. The upper plot shows the classification with a single threshold, the bottom plot uses several different thresholds to optimize the amount of FP and FN as well as the TP and TN. Note that, each threshold was optimized using accuracy (see Chapter 2.3 Equation 2.4) as a metric.

While this plot is an artificially generated example, it is possible to find similar real world scenarios such as the different consumption patterns on weekends and weekdays or clusters of different smart meters with similar consume, which can benefit from different statistical models. Naturally, smaller classes also have a smaller sd, which is noise.

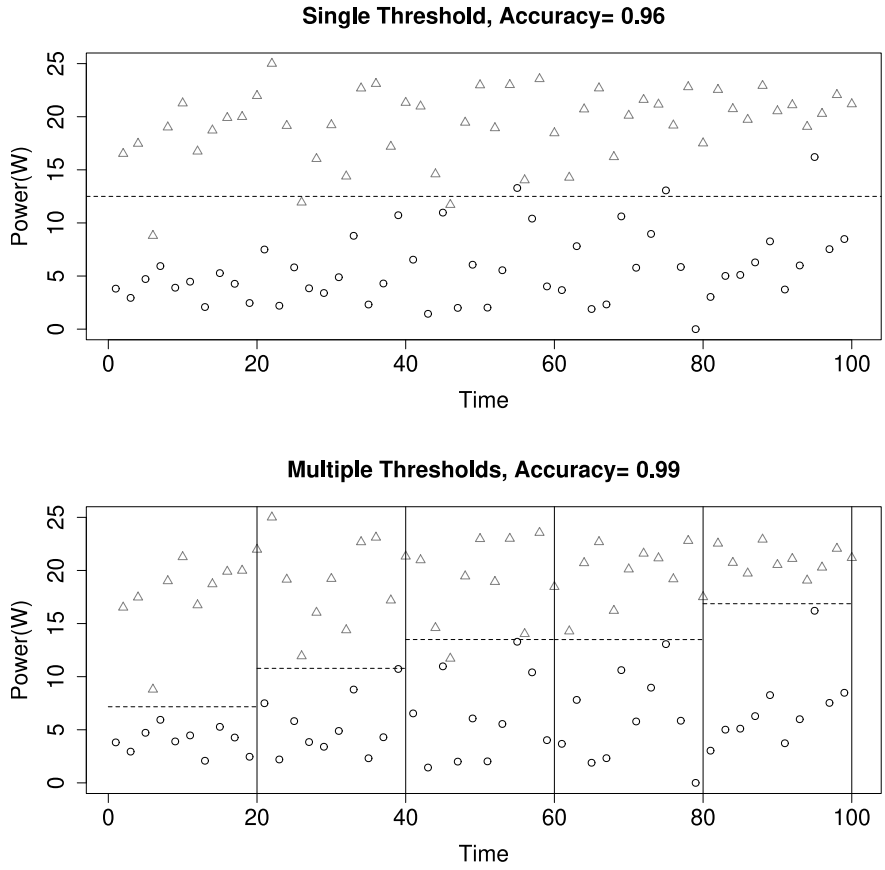


Figure 6.1: Classification of data with multiple thresholds.

The optimal case would be statistically independent groups: when a and b denote any two time series where a and b are independent, then $var(a + b) = var(a) + var(b)$.

For the following experiments, three dimensions are addressed, including m smart meters with n measurements of k days. As a result, the dimensions in the matrix compare:

1. Time: n measurements for adjacent times (traditional)

2. Households: measurement at time i for m smart meters

3. Dates: measurement at time i for k different periods of time

The following example, where a single dimension is insufficient for detection, demonstrates the utility of different dimensions. In this context, the concept of anomalous is defined as an outlier different from the majority of compared data. Let's assume a $n \times m \times k$ matrix $\mathcal{M}$, with identical values x, with some of these values being manipulated through division (identified in grey in the equation below):

$$\mathcal{M}^1 = \left[\begin{bmatrix} x_{111} & \cdots & x_{1m1} \\ \vdots & \ddots & \vdots \\ x_{n11} & \cdots & x_{nm1} \end{bmatrix} \begin{bmatrix} \cdots \end{bmatrix} \begin{bmatrix} x_{11k} & \cdots & x_{1mk} \\ \vdots & \ddots & \vdots \\ x_{n1k} & \cdots & x_{nmk} \end{bmatrix} \right] \tag{6.2}$$

By aiming to detect outliers different from the majority, one can see that the result greatly depends upon the dimension: comparing $x_{111}, x_{211}, ..., x_{n11}$ (in dimension 'time') results in no detection because all values are equally manipulated. Comparing $x_{111}, x_{121}, ..., x_{1m1}$ (in dimension 'date'), half of the values are manipulated and cannot distinguish normal and manipulated values. Comparing $x_{111}, x_{112}, ..., x_{11k}$ (in dimension 'households') an outlier can be seen, because only x_{111} is

[1] The author simplified the matrix notation of each row and column $x_{111}, x_{121}, \cdots, x_{1m1}$ for space considerations.

different from all other values.

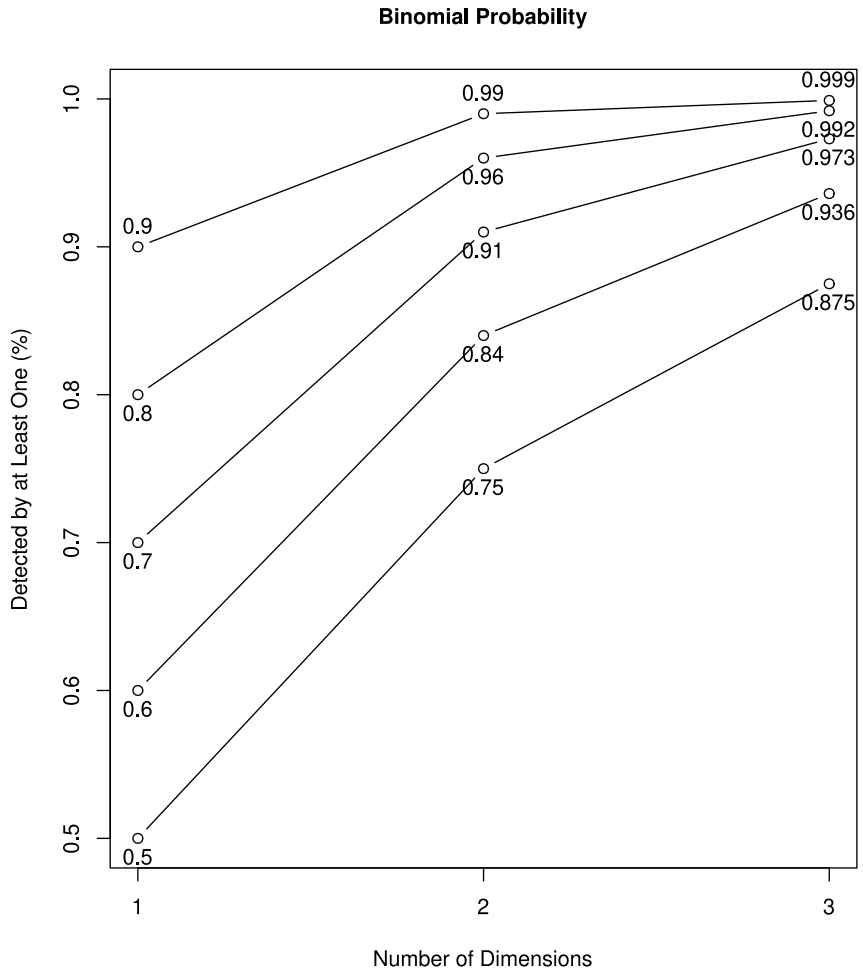


Figure 6.2: Binomial probability for different dimensions.

The relation of dimensions and detection rate can be simplified by considering each dimension of the matrix as individual anomaly detection approach. By assuming that each dimension is statistically independent and used for classification simultaneously, the binomial formula $P = \binom{n}{k} \cdot p^k(1-p)^{n-k}$ can be used to compute the total probability for

detection. Whereas P computes the cumulative probability that the anomalous value is k times successfully detected in a total of n dimensions, if each dimension has the same individual detection rate of p.

Figure 6.2 visualizes the probability that at least one dimension detects the manipulated data. Each line shows a different probability for a correct classification of the individual dimensions, whereas the x-axis shows the amount of trials. The y-axis shows the cumulative probability for at least one correct classification.

6.2 Anomaly Detection using Multiple Dimensions

Due to the daily pattern of energy load curves, the detection of relevant outliers by comparing different data sources (e.g. electricity meters) or different datasets of one data source (e.g. previous days) can be a challenging task. This section introduces and validates requirements on the input data of the entropy-inspired outlier preserving metric and highlights cases in which the metric leads to a concrete improvement compared to alternative methods aggregating different data sources. Here, the previous two scenarios to tamper with electricity meter data are utilized for energy theft (see Remark 5.3.1), which leads to noticeable outliers. The chapter first evaluates the accuracy of the method using the ECO data and subsequently compares the method to other anomaly

detection schemes used on energy demand.

For the approach, the household characteristics introduced in Chapter 5 are used. These values, which are the input data for the aggregation method, are referred to as 'feature' to separate them from the 'metric', which is the output of the entropy-inspired aggregation method. These features build the underlying statistical model and distinguish between regular and anomalous behaviour. In the proposed method, each feature is compared to another data source, such as a previous day of the same household or a household with similar energy consumption: the features are aggregated by encoding their distribution to a normalized floating point value with the so called entropy-inspired metric, which preserves outliers in the distribution of input features. The resulting time series can be combined with off-the-shelf forecast algorithms, such as Holt Winters, to remove the daily patterns by subtracting the forecast. Now, a simple threshold can distinguish benign and anomalous values.

Figure 6.3 shows the complete process to build the normal model and subsequently detect anomalous time windows. The left part illustrates the data at each step of the process: the top shows the raw input of different sources as line chart, the middle shows the derived features, whereas each time window results in a stacked bar of all data sources, the bottom shows the resulting entropy-inspired metric. The right part

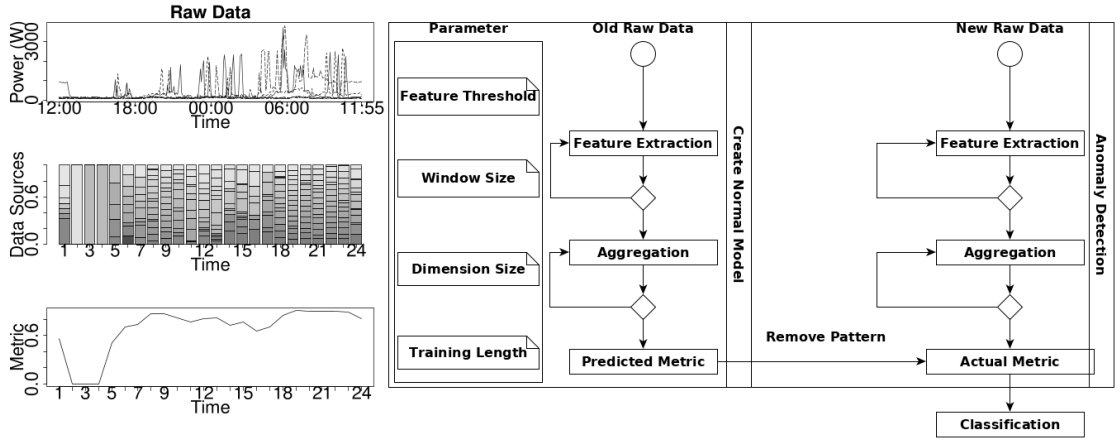


Figure 6.3: Workflow of the anomaly detection theme.

shows training and anomaly detection process with corresponding parameter. While the computation of the metric is equal for both, as last step the forecast model is subtracted from the actual metric to remove the daily pattern. Now, a simple threshold can classify benign and manipulated time windows.

Aside from the previously introduced features the entropy, which encodes the distribution to a single floating point value, is essential for the introduced method. The entropy is a convenient way to aggregate several values without losing information on outliers. The method aims to aggregate the distribution of each dimension to a single value, so that it is possible to apply time series prediction, and hence remove predictable patterns. The reasoning here is that similar data sources, such as two spatially close and structurally identical photovoltaic cells or two days of the same household, should result in similar volumes of features, and

hence a uniform distribution (or at least a consistent pattern). If we compare e.g. different households it is difficult to define a fixed degree of skewness as anomalous. Hence, we measure the entropy regularly to create a time-series of values and a distribution different to the predicted time-series of entropy values is detect as anomalous.

Here, the entropy is not used in a traditional sense, but as a metric to mathematically trace outliers in the distribution of features over time. While the proposed entropy-inspired metric may have characteristics similar to Shannon's entropy, this thesis does not aim to proof that the entropy is the only function fulfilling the requirement to preserve outliers. The practical function as a metric, which encodes distributions without losing information on outliers, is the most important aspect for this work.

Remark 6.2.1. *Formally, let* $\mathcal{X} = x_1, x_2, ..., x_n$*, denote the frequencies of outcomes from the random variable, whereas* $m = \sum_{i=1}^{n} x_i$ *implies the number of all experiments. Then, for a better comparison, the entropy is normalized (as introduced in chapter 3.2) to [0,1] as follows:*

$$\hat{H}(\mathcal{X}) = -\sum_{i=1}^{n} \frac{\frac{x_i}{m} \cdot \log\left(\frac{x_i}{m}\right)}{\log\left(n\right)} \tag{6.3}$$

The input for the entropy could principally be an artificial histogram, such as the relative volume of energy demand. However, if the range of input values changes over time, the entropy is difficult to compare with a previously computed entropy. Therefore, the volume of each feature should have a fixed range. Furthermore, the entropy is profoundly affected by the number of possible outcomes, which means more input values (dimensions) result in the entropy-inspired metric being less affected by outliers.

Let's assume that the initial state is a vector with uniform distribution $v = (\frac{1}{n}, \frac{1}{n}, ..., \frac{1}{n})$ of length n. It is reasonable that a change (a) to $v = (0, 0, 0, 1)$ is more visible than a change (b) to $v = (\frac{1}{n-1}, \frac{1}{n-1}, ..., \frac{1}{n-1}, 0)$, because the sum of change in (a) is $\frac{2(n-1)}{n}$, whereas the sum of change in (b) is $\frac{2}{n}$ (see also Chapter 4 Figure 4.1). According to this reasoning, one can construct the proposed method to detect one of two situations using the entropy metric: detect few falsified value which must be much bigger than the normal values, or detect a majority of falsified values which must be much smaller than normal values. Here, the author prefers the first option in order to detect anomalies early on and tolerate that visibility decreases with increasing proportion of tampered values.

The Figure 6.4 illustrates the entropy of a distribution from real world data. On the x-axis of the plot six time windows can be seen: (a)

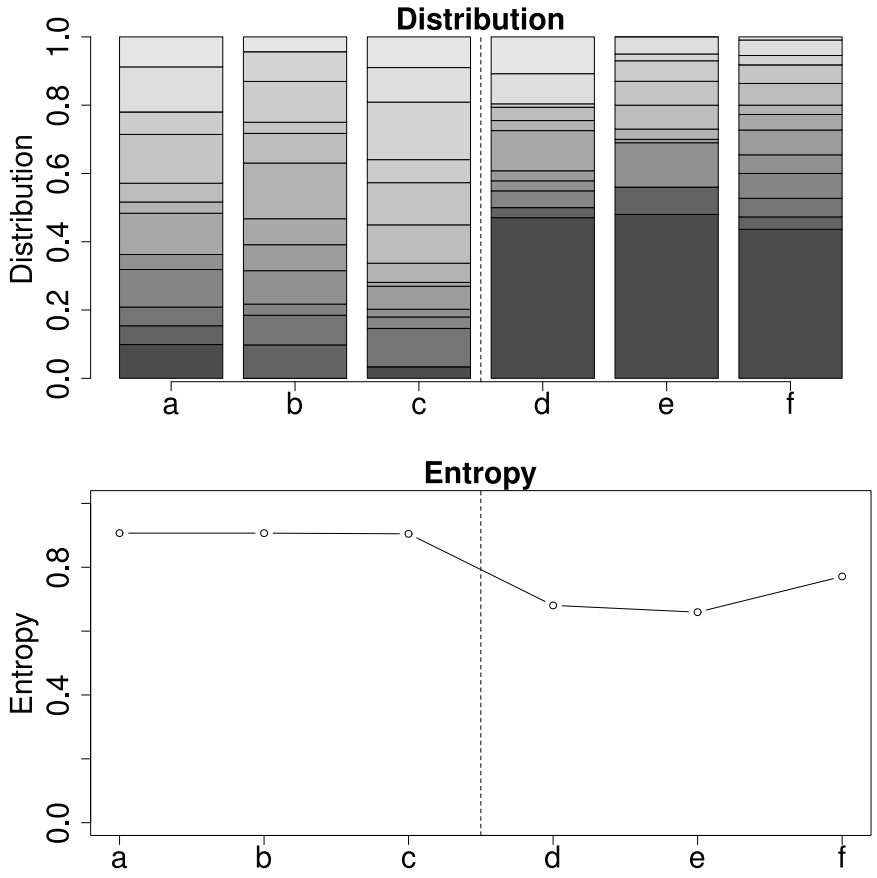


Figure 6.4: Entropy values for energy theft and benign demand.

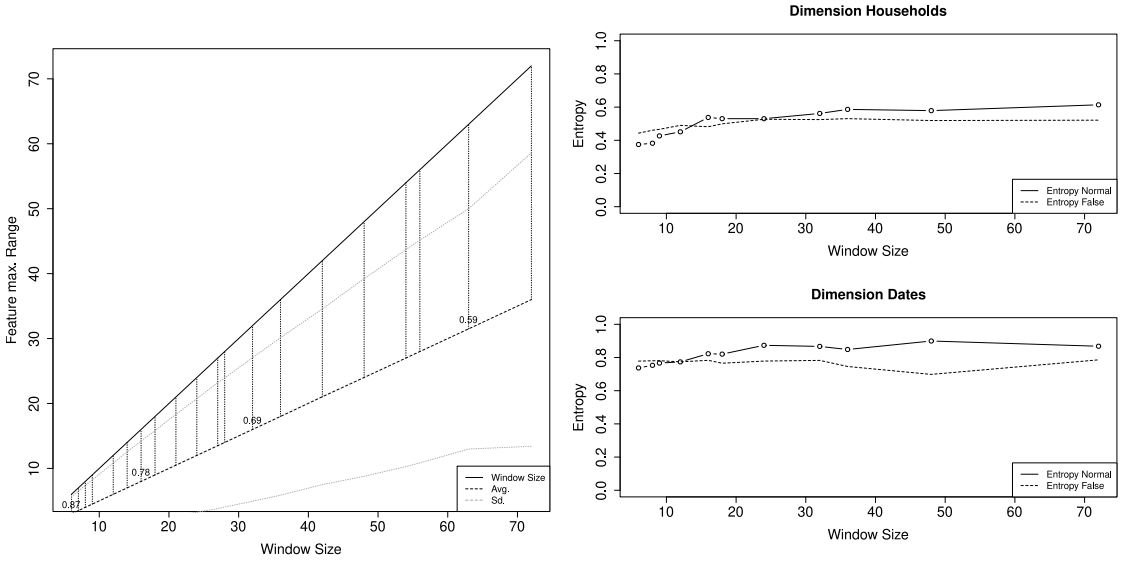
morning, (b) afternoon and (c) evening of sample one, as well as (d) morning, (e) afternoon and (f) evening of sample two with energy theft. In the upper stacked bar plot, each stack (a-f) has twelve elements. Each element represents the size of the feature (f_1: number of high amplitudes) for a different day during the corresponding time window. The left side (a,b,c) with normal energy demand is roughly uniformly distributed, whereas the right side (d,e,f) with energy theft includes an outlier. The bottom line diagram shows the corresponding entropy,

which is high for a uniform distribution and low for a skewed distribution as in case of energy theft.

With the above characteristics of the metric, an anomaly is revealed by very small entropy values. However, such anomalies can also be detected by the outlier in the distribution of input values. Hence, it is always possible to detect the same anomaly by looking at the input values instead of the entropy. The entropy is only a convenient way to aggregate several values without losing this information (e.g. compare several households simultaneously over time). The main motivation to aggregate the results is to simultaneously apply off-the-shelf time series algorithms such as Holt Winters on several households or days, which is difficult with a matrix representation. To apply the entropy horizontally on the dimension 'time' is not suitable as it would mean to encode an outlier which is already visible with other close values and worsen the result.

6.3 Experimental Evaluation

Before the evaluation of the overall quality of the anomaly detection, the next section briefly analyses the parameter which influence the entropy. The change of the entropy can be maximized using two parameters: first, the window size of the feature, which is defined by the amount of mea-



(a) Entropy for different window sizes.

(b) Entropy for normal values only and 1 of 6 samples falsified.

Figure 6.5: Effect of the feature window size on the entropy.

surements in a window, and affects the maximum difference of normal data and outlier. Second, the size of the vector, which is defined by the number of dimensions, and affects the proportion of normal and falsified values in the distribution. Here, these two parameters are examined in conjunction with the entropy.

The window size, which is used to compute a feature, defines the amount of classified measurements, and hence the range of the feature.

Figure 6.5a shows the window size (x-axis) and corresponding avg. feature volume (y-axis) of normal data (mean of 24 days). While the avg. volume is consistent about half of the window size, it can be seen that the sd slowly decreases with the window size (total range) of the

feature. However, a bigger window size can be a disadvantage, because it will only appear as anomalous if the majority of measurements within this window are falsified. Figure 6.5b, shows the average entropy for six normal samples (black) and one of six samples falsified (dotted) for dimension 'household' and 'date'. In this chapter energy theft should always result in an outlier, and hence a small entropy. One can see here, that the distance of the entropy with normal and anomalous data is increasing with the window size of the feature. The avg. entropy is not expected to be a particular good indicator, because not every time window is expected to be uniformly distributed, and hence the avg. entropy is influenced by regular daily pattern. However, the fact that a small window size can result in a completely wrong model, where the avg. entropy of tampered data is greater than the entropy of normal data, indicate that the entropy of such time windows may be hard to predict.

The second parameter needed to address is the 'dimension size', which is of paramount importance to the entropy-inspired metric, because it reflects the necessary distance to other (normal) values in order to appear as outlier. It may sound intuitive that smaller dimension sizes are better, because in a vector of smaller length, individual false values appear proportionally bigger and result in a more skewed distribution which

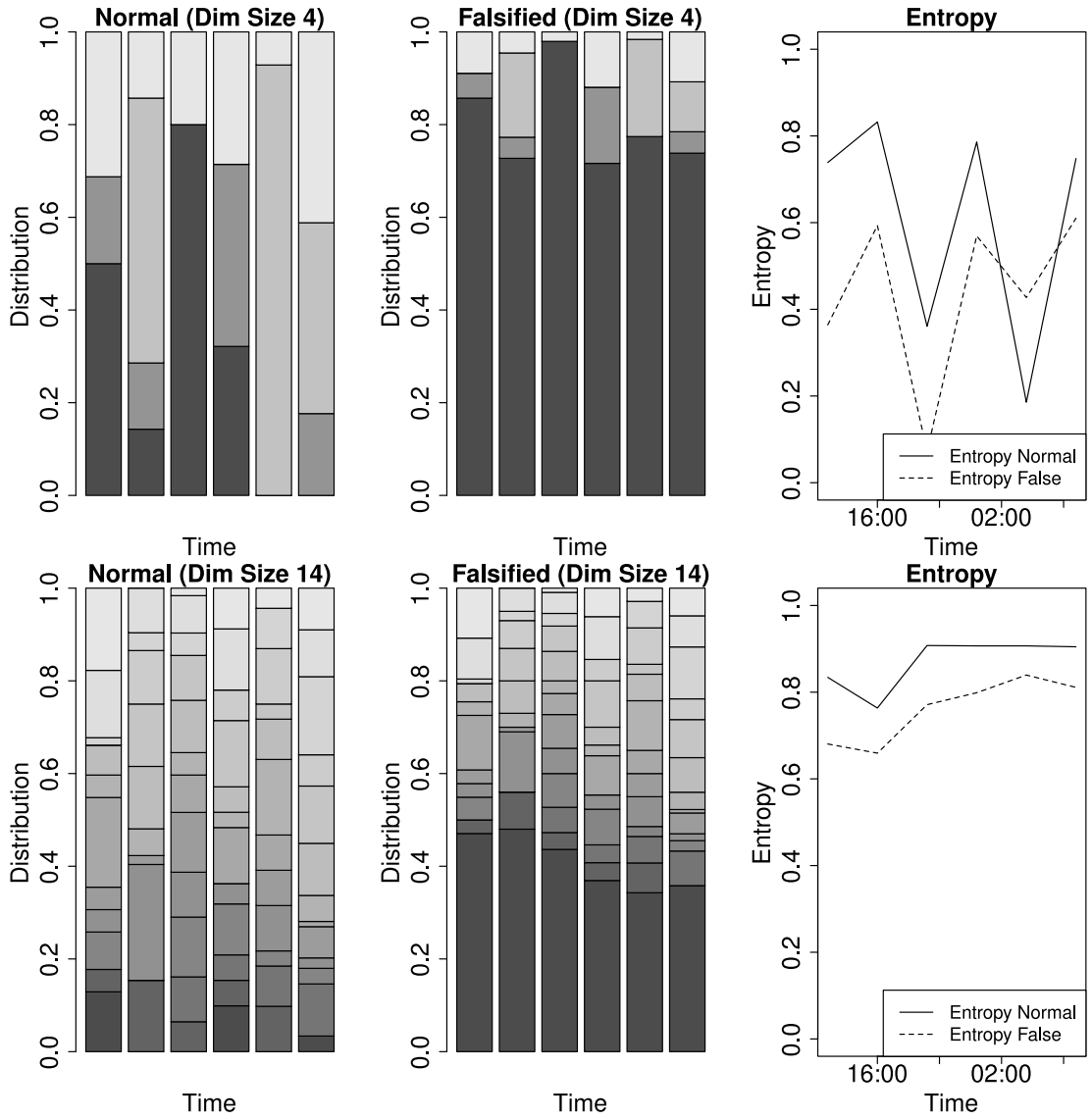


Figure 6.6: Distribution and entropy for two vector sizes.

affects the entropy. But smaller vectors are often heavily influenced by the sd and appear anomalous even without falsified data. Figure 6.6 illustrates this effect by the distribution (bar plot) and entropy (line chart) of the feature over different days in 4 hour windows (48 measurements per window) for two different dimension sizes (top: 4 days;

bottom: 14 days). The left figures show regular days while the middle figures show a sample with one day falsified (type 1). Note that, the author had to look for a specific day with irregular load curve in the dataset to demonstrate this.

According to this reasoning, the optimal dimension size is the smallest possible size, where the method can still consistently distinguish the entropy of anomalous and normal values. This is evaluated using the AUC of a Receiver Operating Characteristic (ROC) curve (see Chapter 2.3).

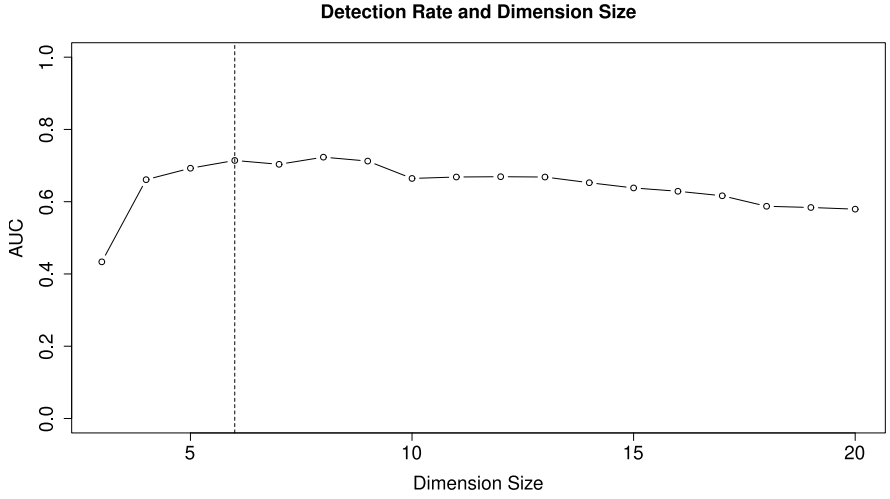


Figure 6.7: Detection rate for different dimension sizes.

Figure 6.7 shows the AUC for different dimension sizes (each dot shows the AUC for 240 classifications = 40 days), which indicates no significant improvement with dimension sizes greater six (vertical dotted line). The AUC still improves with more dimensions if the prediction

is included. The dimension size of six was still used due to practical reasons.

Even in case of good conditions the difference of benign and tampered data is often smaller than the expected sd of the entropy, which is due to the regular patterns. Hence, in order to detect anomalies more reliably, one needs to remove the usual time-depending patterns carried by the features and the resulting entropy. In the following, a time series prediction algorithm is used to forecast the expected entropy. By subtracting the predicted from the actual value the method aims to receive a straight line, without time-depended pattern, which is high for normal and low for falsified data.

For the experiments, Holt Winters, which models the level, seasonality and slope of a time series using training data was employed. A small amount of training data is preferable as privacy concerns and performance may not allow huge sets of historic data for ex post analysis. However, small amounts of training data can lead to over-fitting: while the Hold-Winters model perfectly fits the training data, the actual data would show unpredicted patterns different from the learned data and not match the model.

The length of training data does depend upon previous parameters, such as the window size and dimension, as those parameters define the

expected frequency of regular patterns. Here, the Root-Mean-Square Error (RMSE) between model and actual data was chosen as a rough estimation of the model quality. The RMSE, which is the square root of the variance of the residuals, shows how close the model fits to the actual data. It is in the same unit as the training data, the normalized entropy with a range [0,1], which means an RMSE of [-1,1] can be expected. Values close to zero indicate better fit, as zero means that there is no variance between training data and model.

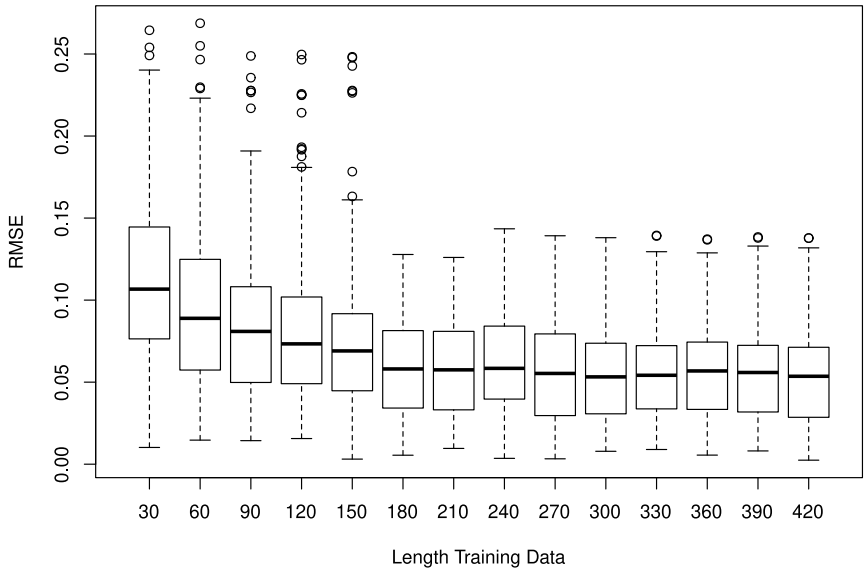


Figure 6.8: Prediction error for different lengths of training data.

Figure 6.8 shows boxplots for the RMSE, ranging from 30 entropy values (5 days) to 420 values (70 days) as training data (avg. of 300 trials). The plot used the feature 'high amplitude' on dimension 'date'

with a window size of 48 power measurements (4 hours) and a dimension size of six to compute the entropy. In this specific case, the RMSE is minimized at 180 entropy values, which is about a month.

Table 6.1: RMSE for each feature and dimension.

H_1	DM	p-value
$f_1 < f_2$	7.7143	4.012e-13
$f_1 < f_3$	7.635	6.393e-13
$f_3 < f_2$	3.6171	0.0001934

Table 6.1 aims to compare which of the introduced features is in general better to predict. For this purpose a Diebold-Mariano test, which is using the forecast error to compare the accuracy of two forecasts with a so called DM statistic, was set up. The DM statistic is used to compute a hypothesis test with the null hypothesis that both forecasts are of equal accuracy, and the alternative hypothesis that one method has greater accuracy. A small p-value indicates strong evidence against the null hypothesis, which means that the null hypothesis can be rejected and the alternative hypothesis is true. A large p-value indicates weak evidence against the null hypothesis, which means that no conclusion can be drawn. Table 6.1 shows the resulting hypothesis test for the number of high amplitude points (f_1), the number of amplitude changes (f_2) and the number of similar amplitudes in a row (f_3), with the same parameters that were used in Figure 6.8.

The advantage of the entropy as a metric is especially the ability to

keep information on outliers after aggregating several values, whereas the aggregate value can be utilized to apply time series prediction. For this reason, the next experiments first compare the introduced metric with an alternative metric and evaluate the value of the time series prediction. In the second part of this section, the complete anomaly detection scheme is compared with other state of the art approaches.

The next question to be asked is whether the time series prediction increases the detection rate in contrast to the same method without prediction. Furthermore, the experiments assess the quality of the entropy to aggregate data while preserving outlier. As an alternative to the entropy, the maximum distance to the mean-value of a row is computed: in contrast to the entropy, this value is large in case of big outliers and small if all values are uniform distributed – this method is called Distance to Maximum (D2M). Given is, for both methods (D2M and Entropy), a time series consisting of individual values, which summarize several dimensions over a predefined time window. Each method is evaluated in combination with Holt Winters (a) and without (b). The decision algorithm for both methods differs slightly: for (a), a value is anomalous if the metric exceeds a threshold. For (b), a value is anomalous if the actual metric minus prediction exceeds a threshold.

The table shows the AUC of both methods with a so called bootstrap

Table 6.2: AUC: type 1 falsified.

		Dimension: Date						
		f_1	f_2	f_3		f_1	f_2	f_3
1 of 6	Ent+TS	0.953	0.951	0.938	Ent	0.76	0.684	0.661
	D2M+TS	1	0.998	0.938	D2M	0.973	0.854	0.761
	p-value	0.014	0.00607	0.99	p-value	4.1e-09	5.98e-10	0.000323
2 of 6	Ent+TS	0.942	0.981	0.907	Ent	0.767	0.771	0.624
	D2M+TS	0.893	0.921	0.742	D2M	0.626	0.675	0.523
	p-value	0.0547	0.0229	5.4e-06	p-value	2.8e-05	7.37e-08	0.333
3 of 6	Ent+TS	0.853	0.975	0.822	Ent	0.646	0.756	0.557
	D2M+TS	0.621	0.853	0.518	D2M	0.657	0.587	0.672
	p-value	1.31e-07	0.00308	3.6e-11	p-value	0.911	1.13e-12	0.251
		Dimension: House						
		f_1	f_2	f_3		f_1	f_2	f_3
1 of 6	Ent+TS	0.456	0.87	0.591	Ent	0.68	0.83	0.507
	D2M+TS	0.689	0.926	0.669	D2M	0.461	0.94	0.633
	p-value	0.0173	0.00105	0.103	p-value	1.74e-08	0.000145	0.206
2 of 6	Ent+TS	0.561	0.956	0.609	Ent	0.729	0.897	0.707
	D2M+TS	0.629	0.956	0.531	D2M	0.818	0.894	0.577
	p-value	0.127	0.944	0.0876	p-value	0.00223	0.894	0.00163
3 of 6	Ent+TS	0.672	0.876	0.682	Ent	0.851	0.815	0.76
	D2M+TS	0.758	0.829	0.589	D2M	0.927	0.731	0.644
	p-value	0.0368	0.0679	0.0137	p-value	0.0029	0.00103	0.00129

test. Here, the AUC is repeatedly ($N = 2000$) computed with the original inputs for the ROC curve re-sampled, which approximately follows a normal distribution used to perform a hypothesis test. The null hypothesis is that the true difference between both AUC is zero and the alternative hypothesis is that method one performs better than method two. A small p-value shows that the null hypothesis can be rejected. Table 6.2, shows the result of this experiment for each feature, the number of high amplitude points (f_1), the number of amplitude changes (f_2) and the number of similar amplitudes in a row (f_3), with dimension size 6 and a window size of 48. Furthermore, the table shows the AUC for

one, two or three corrupted samples (days). Note that, a majority of falsified samples cannot be detected per definition. To train Holt Winters, 6 entropy values (1 day) were predicted from the last 60 entropy values (2 month). Each experiment was repeated (over a time of 4 month) to get 120 prediction values, which are used to construct the AUC of a ROC Curve. In the results, it can be seen, that the time series prediction significantly improves the results for both metrics, the entropy and D2M. Although the entropy-inspired metric, in combination with time series prediction, results in acceptable detection rates mostly over 90%, which can compete with the other function, it only performs better with several outliers in the distribution – which means that e.g. the current day and day before show energy theft. A case with several outliers is difficult to detect for the alternative function.

The lower part of the table shows the same results for dimension 'household'. Unfortunately, due to the different load pattern in the dimension 'household', two of three proposed features were not working well enough for a practical usage. For features 'high amplitude' and 'similar in a row', it was not possible to optimize the parameter ε well enough to get an approximately uniform or otherwise predictable pattern for each household. However, the feature 'amplitude changes' (f_2), was resilient to these different load patterns and performed very well (This

is not surprising, because the previous experiments already pointed out that the *côrr* and RMSE perform best for feature 'amplitude changes').

Table 6.3: AUC: type 2 falsified.

		Dimension: Date						
		f_1	f_2	f_3		f_1	f_2	f_3
1 of 6	Ent+TS	0.912	0.792	0.819	Ent	0.73	0.508	0.497
	D2M+TS	0.998	0.852	0.838	D2M	0.969	0.654	0.556
	p-value	0.000932	0.0713	0.637	p-value	6.47e-10	2.04e-05	0.0272
2 of 6	Ent+TS	0.939	0.879	0.673	Ent	0.749	0.378	0.449
	D2M+TS	0.936	0.846	0.584	D2M	0.645	0.404	0.601
	p-value	0.857	0.386	0.0128	p-value	0.00204	0.139	0.14
3 of 6	Ent+TS	0.847	0.843	0.65	Ent	0.644	0.607	0.4
	D2M+TS	0.686	0.797	0.545	D2M	0.629	0.501	0.677
	p-value	8.4e-05	0.102	0.294	p-value	0.874	0.319	0.00517
		Dimension: House						
		f_1	f_2	f_3		f_1	f_2	f_3
1 of 6	Ent+TS	0.456	0.87	0.591	Ent	0.68	0.83	0.507
	D2M+TS	0.689	0.926	0.669	D2M	0.461	0.94	0.633
	p-value	0.0173	0.00122	0.102	p-value	1.47e-08	7.89e-05	0.206
2 of 6	Ent+TS	0.561	0.956	0.609	Ent	0.729	0.897	0.707
	D2M+TS	0.629	0.956	0.531	D2M	0.818	0.894	0.577
	p-value	0.126	0.944	0.0701	p-value	0.00165	0.895	0.00127
3 of 6	Ent+TS	0.672	0.876	0.682	Ent	0.851	0.815	0.76
	D2M+TS	0.758	0.829	0.589	D2M	0.927	0.731	0.644
	p-value	0.0327	0.0755	0.0153	p-value	0.00274	0.000817	0.0016

The results for type 2 falsified data in Table 6.3 are similar. It was expected that some features react better to tampering on the amplitude and others on changed load patterns, but that was not the case. All features react very well on changed load patterns (type 1) and worse if the amplitude not affects load patterns (type 2). 'amplitude changes' performed well, while both other features did not detect energy theft in this case and can only work if the load patterns of the data sources are very similar. The author believes that the performance may be further

increased by clustering similar load curves (in order to reduce the complexity of daily patterns) – which is not in the scope of this work. The feature 'amplitude changes' showed results without clustering households according to their similarity and performed well solely because the pattern of each individual household was consistent enough. A generalization to other datasets is difficult, but the experiments showed evidence that the concept works on the condition to find a feature with consistent pattern on each data source.

The next section introduces a final comparison of the proposed scheme with two other state of the art anomaly detection methods on energy demand, namely a method inspired by AMIDS from McLaughlin, Holbert, Zonouz, and Berthier (2012), which models the energy consumption behaviour of a household using Naive Bayes, and a method based on XMR charts from Spirić et al. (2015). Mclaughlin's original article utilizes NIALM profiles to associate each on/off amplitude in a households energy load curve to a certain appliance. The three resulting vectors with amplitudes, appliance names and on/off operations are used as input for the supervised learning of the Naive Bayes algorithm, which computes the probability for energy theft for each data point. In contrast to the previous method, AMIDS has a strict requirement for high data resolution. If the amplitudes of individual devices are not visible,

the detection rate is highly decreased. For the following experiment, it was necessary to simplify Mclaughlin's method, because the setup of a suitable NIALM database is beyond the scope of this work. Mclaughlin evaluated his method with an energy demand simulation, where the mapping of appliances to predefined profiles is generally easier. The experiment in this work used a simple clustering algorithm instead of an elaborated, hand-labelled NIALM database. Hence, the method assumes that the most frequent amplitudes (as arranged by the cluster centres) correspond to the amplitudes of different devices. One of the limitations of such a method is that appliances with similar amplitude and appliances used together cannot be recognized as individual device. The experiments indicate that the results are still consistent enough for anomaly detection since they resulted in a detection rate similar to Mclaughlin's. However, it may be possible to further increase the detection rate of this method by using a better NIALM algorithm and higher resolution data. Spiric's fraud detection is based on monitoring the 'random component' of the energy demand, which means that the raw input data is decomposed into a seasonal component, trend component and random component (e.g. by using a moving average time series decomposition algorithm). In order to define the threshold for energy theft, Spiric utilizes a so called XMR chart, which computes an upper

and lower limit using the mean moving range. Note that, the threshold of the XMR chart is not relevant for the results, because the AUC was used as metric. The AUC computes the result for any possible threshold, which means that the here computed results of Spiric's method may be slightly better than results with a fixed threshold.

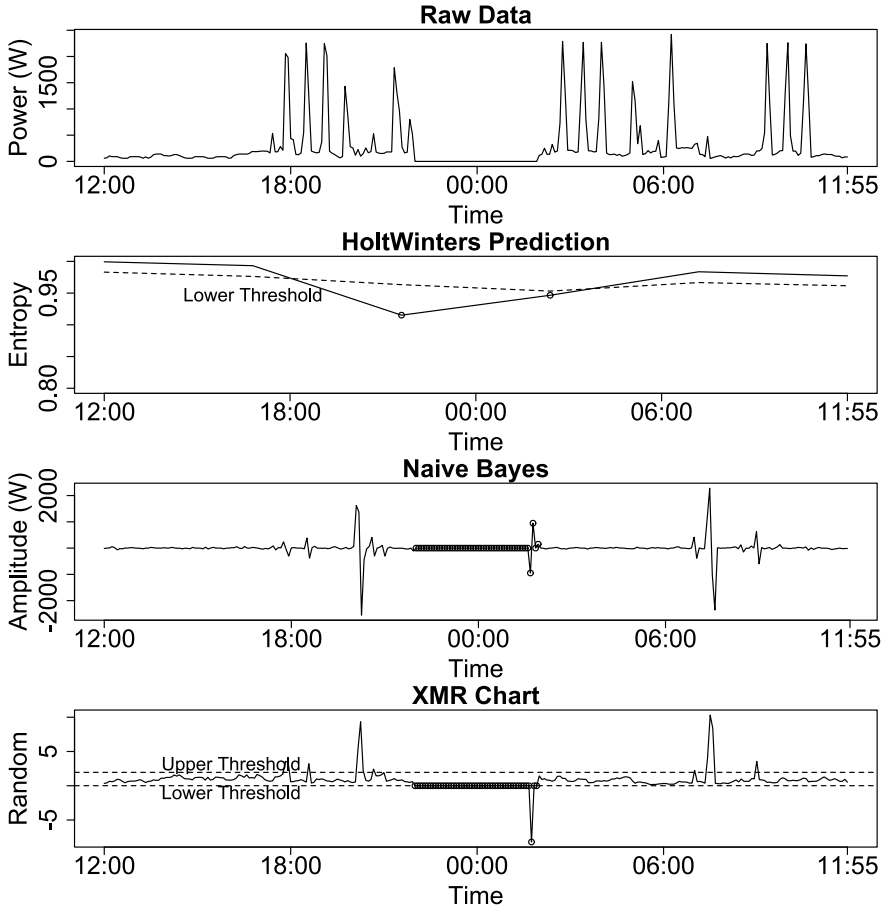


Figure 6.9: Overview of anomaly detection methods.

Figure 6.9 shows an overview of these methods: the top plot shows the raw input data with four hours of energy theft around 0:00 pm. The next plot show the entropy-inspired method, which computes the

entropy (black line) and the Holt Winters prediction of the expected entropy (dashed line) as a lower threshold. Here, energy theft results in a small entropy. The third plot shows Mclaughlin's method, which is using a time series of amplitudes as input data and assigns a probability for energy theft, which is the output of Naive Bayes, to each measurement. Here, energy theft results in a high probability. The bottom plot shows the XMR chart with the random component of time series decomposition as input, which detects energy theft with a lower threshold (dashed line). Here, energy theft results in a small or negative number relative to the mean of the random component.

In order to compare these different methods, it was necessary to accept some limitations. E.g. it may be possible that, especially the Naive Bayes method, can be improved with higher resolution data, because the edges of on/off operations are more visible. Furthermore, it may be possible to optimize the lengths of input data, e.g. the amount of training data or length of the expected seasonality for the decomposition algorithm. For the Naive Bayes the experiments used the previous day as training data, because using more training data worsened the results. For the decomposition, the default setting of one day seasonality was used.

Since the entropy-inspired method has a low resolution output data,

it was necessary to aggregate the output of both other methods to the same resolution in order to compute a consistent AUC. Both methods utilize only a single source of input data, and hence it is not possible to check the results of multiple compromised sources, which is one of the strong points of the entropy-inspired algorithm. Since the previous experiments already contains a detailed evaluation of the entropy-inspired method (see Table 6.2), this experiment was only conducted for the feature 'amplitude changes'.

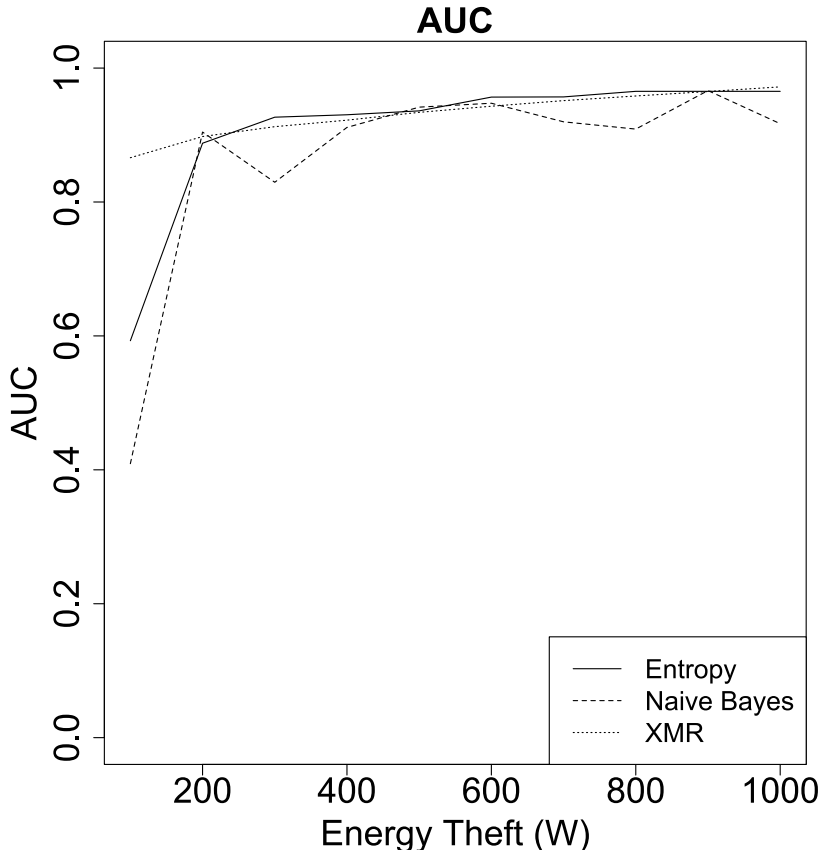


Figure 6.10: Detection rate depending on the amount of energy theft.

Figure 6.10 shows the results for different amounts of energy theft.

Each algorithm aimed to detect one day of energy theft on ECO data household 1 in August. The plot shows the mean of 15 experiments. The output of Naive Bayes still has a lot of variance because the result depends on the random cluster centres used to determine the appliances. The x-axis shows the AUC and the y-axis shows the amount of power subtracted from the original energy demand. It can be seen that Spiric's method is especially good at detecting smaller amounts of energy theft.

Table 6.4: AUC: XMR charts, naive Bayes and entropy.

	AUC		AUC		AUC
EntTS	0.965	EntTS	0.965	Naive	0.962
Naive	0.962	XMR	0.972	XMR	0.972
p-value	0.859	p-value	0.712	p-value	0.475

Table 6.4 shows again the bootstrap test with the alternative hypothesis that the accuracy of both methods is different. However, for the type 1 energy theft, it is not possible to clearly reject the null hypothesis that all methods perform equally good, as all methods have a high accuracy. As the introduced method is generally intended to use additional data sources, which are not considered in the other two methods, these three methods should be used together to complement each other.

6.4 Discussion

This chapter showcased anomaly detection in different dimensions, which can unveil, otherwise hidden, anomalous data if the majority of data in

a single source is compromised. The focus was an aggregation method, which preserves outliers, to remove repeating patterns. The experimental evaluation analysed parameters of the entropy, such as the optimal window size to minimize the sd of normal data. Furthermore, the chapter evaluated the influence of the number of data sources, so that, after the aggregation, energy theft still produces outliers which are greater than the sd of normal data. The chapter demonstrated how to find the optimal length of training data to maximize the predictability of the metric and remove the repeating pattern from the aggregated data. A limitation specific to this anomaly detection method is the low output resolution, which was required to reduce the sd of the feature. Colloquially speaking, it means that the majority of the electricity in this time window must be manipulated for a detection. To investigate RQ5, the introduced approach with multiple dimensions was compared to traditional approaches. While other approaches can reach similar detection rates, the author argues that both approaches together can find otherwise undetected outliers, and hence multiple dimensions can improve the detection rate. For the two scenarios of energy theft, the experiments resulted in detection rates above 90%, whereas the number of amplitude changes $\geq \varepsilon$ performed especially well. The method was able to detect tampered data even by utilizing different households, which were not

clustered according to their similarity, as data source. Removing daily patterns with Holt Winters significantly improved the detection rate from about 75% to above 90%. Apart from the entropy-inspired metric, other aggregate methods may work as well, but the entropy-inspired metric is especially robust in presence of multiple outliers. The detection rate of the alternative aggregation method (D2M) decreased up to 10% for each additional compromised data source, while the detection rate of the entropy-inspired metric did not significantly drop with up to half of all data sources compromised. Sophisticated and stealthy tampering methods, were not analysed in this context, as the detection of data mimicking attacks is a well-known challenge and inherent limitation of anomaly detection and beyond the scope of this work.

Chapter 7

Analysis of Stealthy Energy Theft

This chapter aims to point out the constraints and limitations of anomaly detection with regard to sophisticated and stealthy energy theft methods. The following experiments are inspired by 'Bouché, J., Hock, D., & Kappes, M. (2016). On the performance of anomaly detection systems uncovering traffic mimicking covert channels. In Proceedings of the 11th international network conference (inc) (pp. 19-24).', whereas Johannes Bouché conducted experiments concerning the manipulation of network traffic for SnortAD and the author of this thesis focused on the statistical analysis and implementation of forecast models. As the original article focused on network traffic, the following sections motivate the relevance and application on energy theft and points out differences particular to energy demand. This chapter investigates RQ6: 'Is it possible to detect stealthy manipulation attempts?'.

7.1 Introduction

The security of digital electricity meters confronts us with several new and challenging problems. Analysing meter data across the entire customer base, to find outstanding events and discover inadvertent or deliberate risks, is a sophisticated data mining challenge. With digital access to measurements and metering information, the adversary can arbitrarily change data, which offers potential for sophisticated and stealthy manipulation scenarios.

Anomaly detection methods aim to tackle these new challenges by modelling load curves to find deviations from the normal behaviour. However, many inherent limitations of anomaly detection, which are well-known in established areas such as network traffic monitoring and network intrusion detection, are still neglected. Many methods, such as sophisticated stealthy manipulations, introduced by Casenove (2015), or tampering with the learning algorithm and compromising the detection system itself, as introduced by Corona et al. (2013), have been proposed to the same extent in these areas. Many of these constraints can be conveyed to load curves, and therefore showcase concept and condition for stealthy load curve manipulations. In particular, this chapter shows that, with sufficient knowledge of underlying detection techniques, it is

possible to compute an error margin and tamper with load curve data in a targeted manner to bypass an anomaly detection systems by mimicking legitimate behaviour.

In the following, it is assumed that an adversary has gained access to a corrupt smart meter, including unrestricted access to measurement data. The following experiments show that stealthy manipulations can perform well enough to perform billing fraud, particularly if the metrics used by the anomaly detection system are known. This chapter demonstrates a proof-of-concept, with two previously introduced anomaly detection algorithms, namely an algorithm based on Holt Winters forecast and a method based on Naive Bayes.

Many recent articles, such as McLaughlin et al. (2013), Mashima and Cárdenas (2012) and Jokar, Arianpoo, and Leung (2015) propose anomaly detection systems to detect electricity theft. Some authors, such as Ashrafuzzaman et al. (2018), Bhattacharjee and Das (2018) and Y. Liu, Liu, Sun, Zhang, and Liu (2020) mentioned stealthy manipulation methods in the different context of false data injection. But to the best of the authors knowledge, a performance evaluation of stealthy manipulation, mimicking normal energy demand, has not been investigated before. Concepts similar to the proposed scenario have been presented by Wendzel and Keller (2014) conducting data manipulation

for covert channels or D. Wagner and Dean (2001) mimicking legitimate behaviour in context of network traffic. Urbina et al. (2016) generalized the concept of stealthy attacks and showed that mimicking attacks can be applied to industrial systems.

The remainder is organized as follows. First, the anomaly detection methods, with detection rates above 90% for traditional tampering methods, are introduced. Then, the limitation of these methods are analysed on load curves with artificially falsified data. The experiments compute manipulated load profiles below the error margins and demonstrate that, the error margin can be minimized by using multiple anomaly detection systems which indeed limits the amount of undetected stolen energy.

7.2 Mimicking Holt Winters and Naive Bayes

In the following, the mechanics of the used anomaly detection model is addressed to derive the concept of stealthy manipulation. For the experimental setup, the ECO dataset is utilized. The concept of a stealthy attack can in general be applied to any anomaly detection method or metric, here two methods based on the previously introduced anomaly detection concepts are utilized. Both methods are simplified in order to focus on the aspect of stealthy manipulation. However, the results

demonstrate that the corresponding method can indeed detect the tampering methods, before the stealthy method is applied, with an accuracy of above 90%.

Holt Winters

Time series approaches, e.g. used by Hock, Kappes, and Ghita (2020) or Spirić et al. (2015) are common to detect anomalies. Here, a well-known anomaly detection system developed by Szmit, Szmit, Adamus, and Bugała (2012) is adapted for network traffic monitoring. The method is exceptionally simple and allows this work to focus on the stealthy manipulation. The original method, called SnortAD, repeatedly counts the number of network packets over a fixed period of time and, with a sufficient amount of training data, aims to predict the future amount of packets. The prediction model results in a so called confidence band, which is a lower and upper threshold of packet numbers and defines the error margin of the model. The confidence band defines a range in which the algorithm is confident that packet numbers are normal. In order to adapt this method to energy demand, the author excluded linear prediction methods such as moving average and auto regression, because energy demand has in contrast to network traffic many legitimate periods without loads, which increase the importance of a seasonal compo-

nent in the prediction method. For simplicity, only the lower threshold, which corresponds to energy theft, is considered.

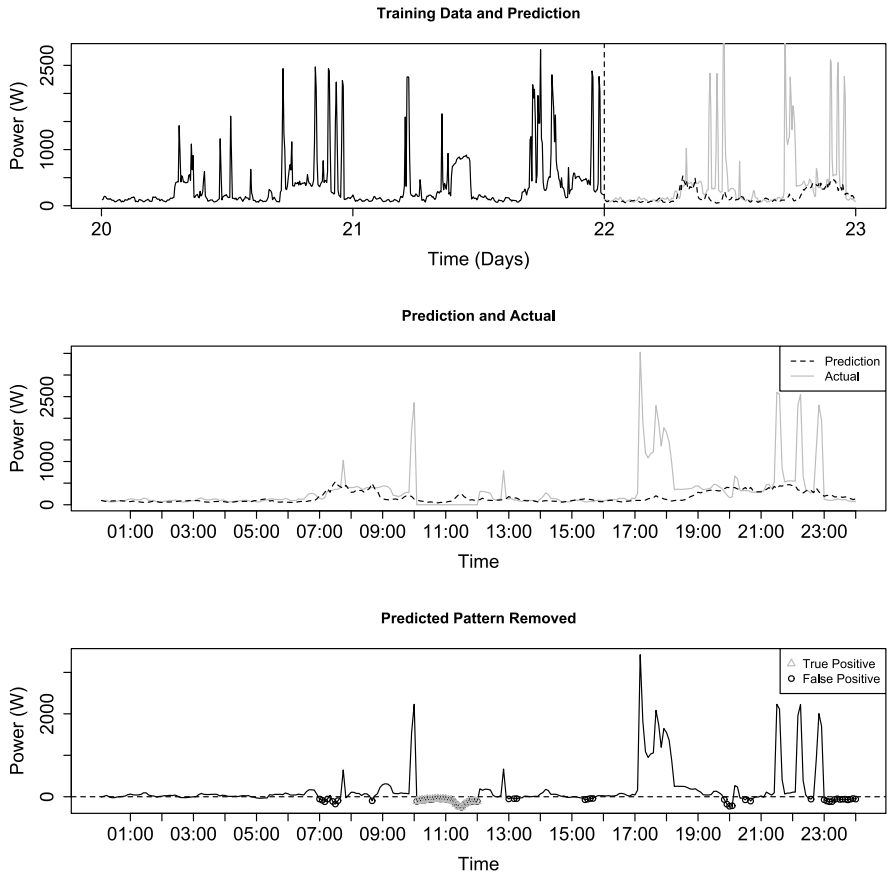


Figure 7.1: Overview of the anomaly detection process.

Figure 7.1 visualizes all three steps of the process, using the Hold Winters model on several weeks of energy demand captured in the ECO data set. The first plot shows the last days of learning data for the Holt Winters Algorithm in black and after the vertical dotted line a comparison of the prediction (dotted) and actual (grey) data. The second plot shows this comparison on a larger scale. Here, it is clearly visible

that a time window (10:00 - 12:00) of the data is manipulated. In the last plot, the prediction is subtracted from the actual data to find data which is unusual low. The grey triangles show the detected manipulation while the black dots visualize legitimate measurements which are wrongly selected.

Remark 7.2.1. *Formally, consider a finite time series* $\mathcal{T} = x_1, x_2 \ldots x_n,\ x_i \in \mathbb{R}_0^+$ *forall* $1 \leq i \leq n$*, with* n *elements – representing energy demand and the training data for the forecast algorithm.* $\mathcal{T}$ *is input of the Holt Winters function hw, which computes a level, trend and seasonality, whereas* $\hat{\mathcal{T}}$ *denotes the forecast with* m *elements. In order to prevent the model from resulting in negative values, the logarithm is applied before the prediction, resulting in* $\hat{\mathcal{T}} = e^{hw(ln(\mathcal{T}))}$*. The decision function* $f : \mathbb{R}_0^+ \mapsto \mathbb{B}$ *labels a measurement as an anomaly, on the condition that the distance of prediction and actual values fall below threshold* ε*:*

$$f(i) = \begin{cases} 1, & \text{if } \mathcal{T}_i - \hat{\mathcal{T}}_i < \varepsilon \\ 0, & \text{otherwise} \end{cases}$$

In order to demonstrate the accuracy of energy theft detection using the load curve prediction of the next day, two metrics are evaluated: the RMSE and AUC in dependence on the length of training data and data resolution. The RMSE, in the top plots of Figure 7.2, shows the difference between model and actual data, whereas the variance of the

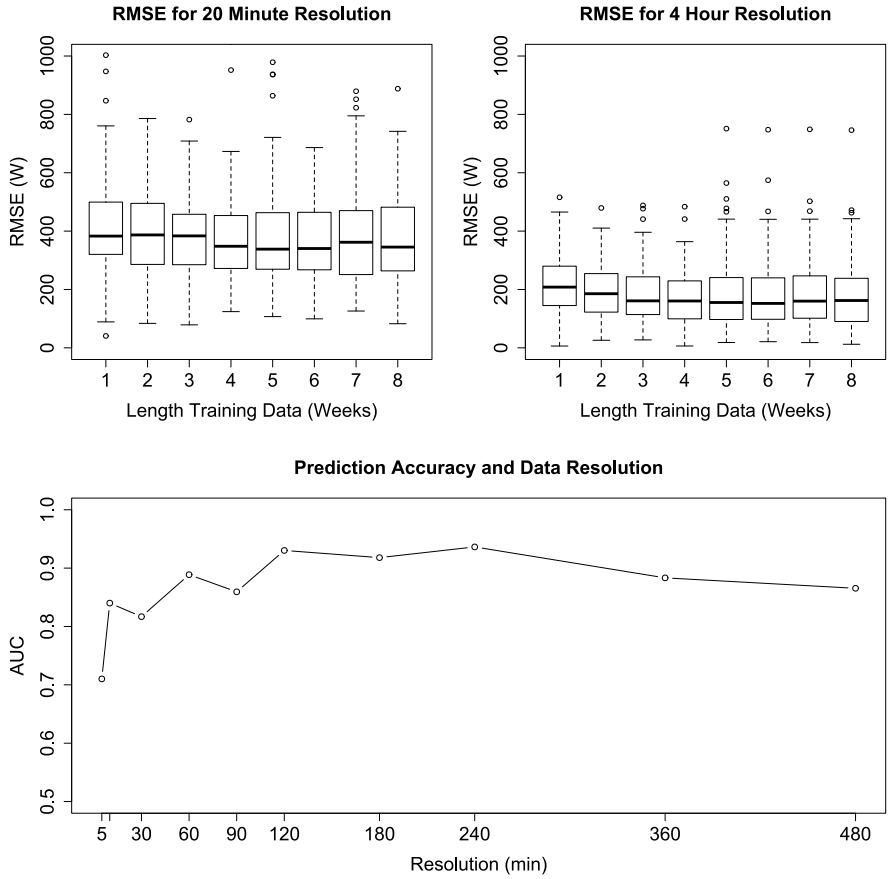


Figure 7.2: Effect of the data resolution.

residuals is the same unit as the training data (W). Here, the actual data was not manipulated, and hence the RMSE quantifies the prediction error of the confidence band and whether values below the band can reliably be detected as outliers. The AUC of a ROC curve, in the bottom plot of Figure 7.2, shows the detection accuracy depending on the data resolution. Here, eight hours of a day are falsified to simulate the disconnection of the electricity meter.

Fine-grained measurements influence the detection accuracy nega-

tively, which is reasonable, because a lower resolution automatically filters high peaks in the energy demand. These peaks cannot be captured by the seasonal component of the Holt Winters prediction, and hence increase the size of the confidence band. However, the smaller prediction error comes with the trade-off that the method can only see manipulation attempts which last this long.

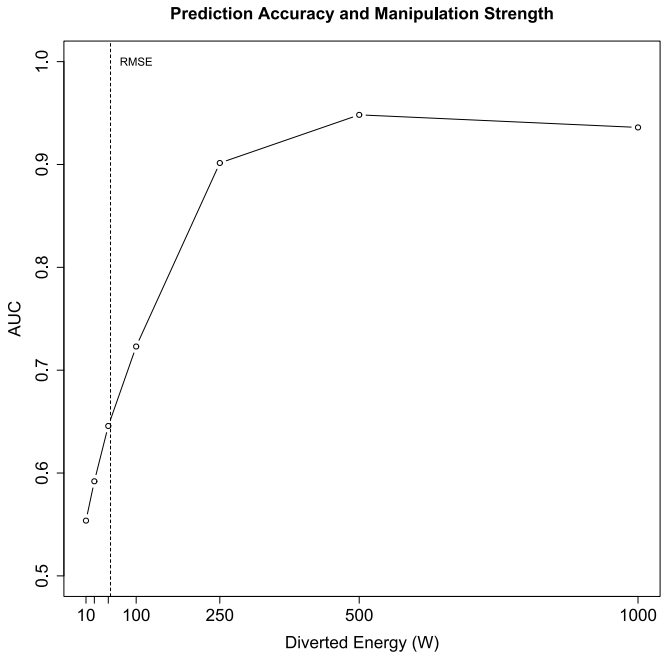


Figure 7.3: Detection rate in contrast to diverted energy (W).

Figure 7.3 shows the AUC (y-axis) for different amounts of energy theft (W) as shown on the x-axis. The experiments used a training length of four weeks and a resolution of one measurement every four hours according to the previous results. The results indicate that, the detection rate is only sufficient with manipulations stronger then the

RMSE (dotted line).

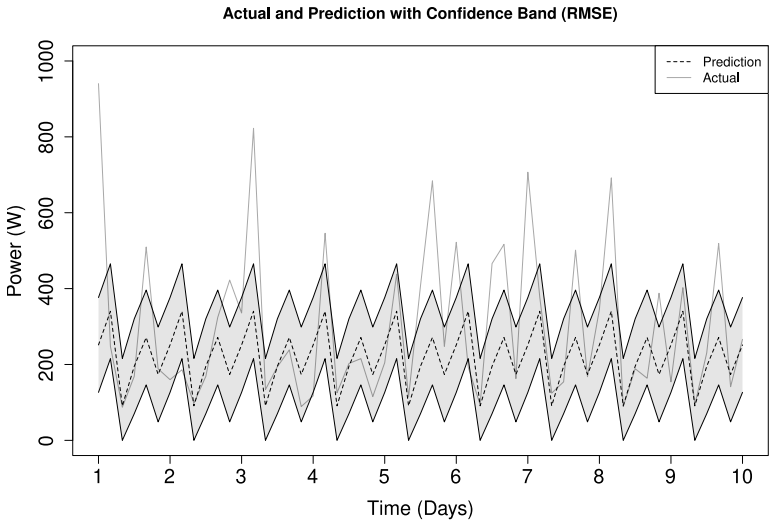


Figure 7.4: Holt Winters prediction with confidence band.

Figure 7.4 visualizes the Holt Winters model with confidence band (Gray area) derived from the RMSE, whereas the lower bound of the confidence band is the minimal energy consumption which is not anomalous. By utilizing this error margin, an adversary attempting to hide electricity theft can follow the expected load closely, while maximizing the power reduction at each time to the point that it does not appear anomalous. Formally, it satisfies the condition $\mathcal{T}_i - \hat{\mathcal{T}}_i \geq \varepsilon$ in the optimal case. Following the above reasoning, the RMSE can be utilized to forge a stealthy energy theft attempt, because the RMSE is the minimal confidence band and any smaller manipulation is difficult to detect.

Naive Bayes

As a second anomaly detection method this chapter uses a method inspired by AMIDS, which models the energy consumption behaviour of a household using Naive Bayes (see Chapter 6.3).

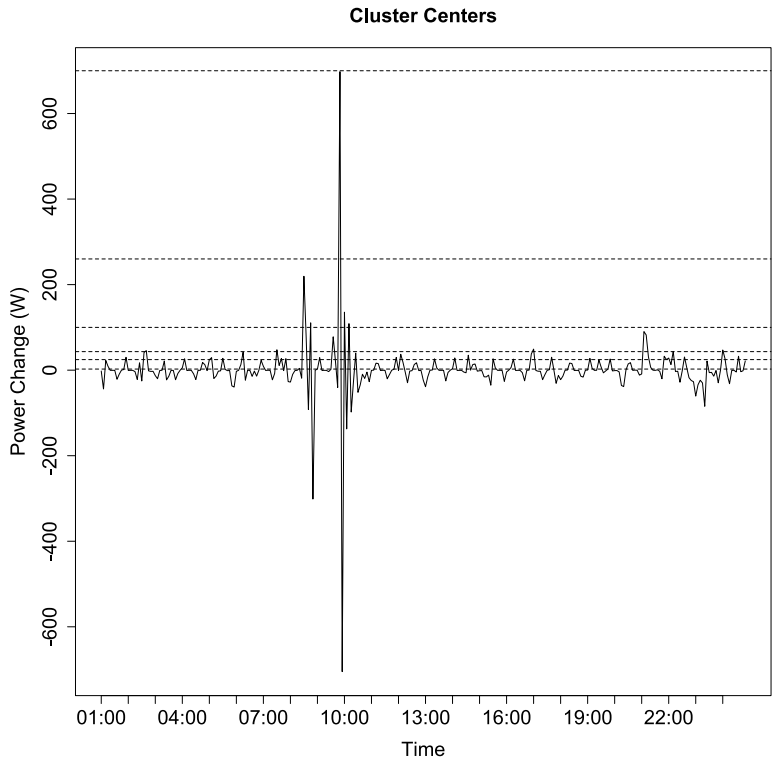
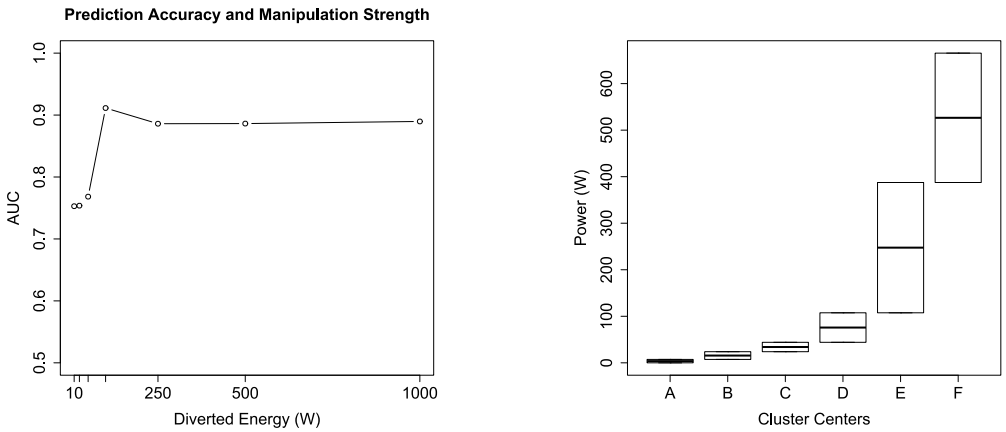


Figure 7.5: Approximation of the device amplitudes.

As in the previous chapter, McLaughin's method was simplified by using k-means clustering instead of using a NIALM algorithm with elaborated and hand-labelled database. Figure 7.5 shows the input data for this process, which is using k-means clustering to capture the most frequent amplitudes. The y-axis shows the difference between the two



(a) Detection rate in contrast to diverted energy (W).

(b) Range of the cluster centres.

Figure 7.6: Anomaly detection with naive Bayes method.

measurements in power over time (x-axis). The horizontal dotted lines visualize the cluster centres which are the approximated devices. Note that, the absolute values of the amplitudes are used, as negative power values correspond to switching of a device off.

Figure 7.6a shows that the detection rate for traditional tampering is, even with this simplification, over 90%. To avoid the additional variance introduced by the random starting point of the k-means clustering, the experiments always show the avg. AUC of ten experiments, which is significant to the third decimal place. In comparison with the previous Holt Winters method, it is noticeable that the detection rate does not necessarily increase with the manipulation strength, even small changes of 10 Watt are still detected with detection rates over 75%, as several cluster centres are located in this region.

Figure 7.6b explains the better detection rate for smaller manipulations. The cluster centres are not uniformly distributed over the min/max range of the power, instead there are several centres in the lower ranges. If the energy demand is manipulated, so that the amount of measurements in a cluster or the sequence of assigned clusters, and hence the input for the Naive Bayes algorithm, is changed then, intuitively, the output of the algorithm changes. However, it can be seen that some cluster centres with higher amplitude cover a greater range of power. By changing each amplitude to the minimal value of the cluster, an adversary can manipulate the energy demand without changing the probability of a transition between devices.

7.3 Experimental Evaluation

This section compares different manipulation methods to evaluate the possible extend of stealthy energy theft and the performance of an intelligent method in comparison to simple methods. The experiments analyse the ratio of anomalous values to energy theft and try to predict the optimal energy theft according to the previously introduced concept. In the experiment the adversary knows method and data, but not the exact parameters such as the length of training data or cluster centres.

Subsequently, four methods to bypass both anomaly detection, start-

ing with the Holt Winters anomaly detection, are compared. Figure 7.7 visualizes these methods, whereas each plot shows a grey shaded area with the actual energy demand (upper) and confidence band of detection (lower). The lower threshold of the grey area is the prediction of the Holt Winters Algorithm of the next seven days (42 Measurements). The black line visualizes the energy theft of each method, whereas measurements below the grey area are detected by the anomaly detection system.

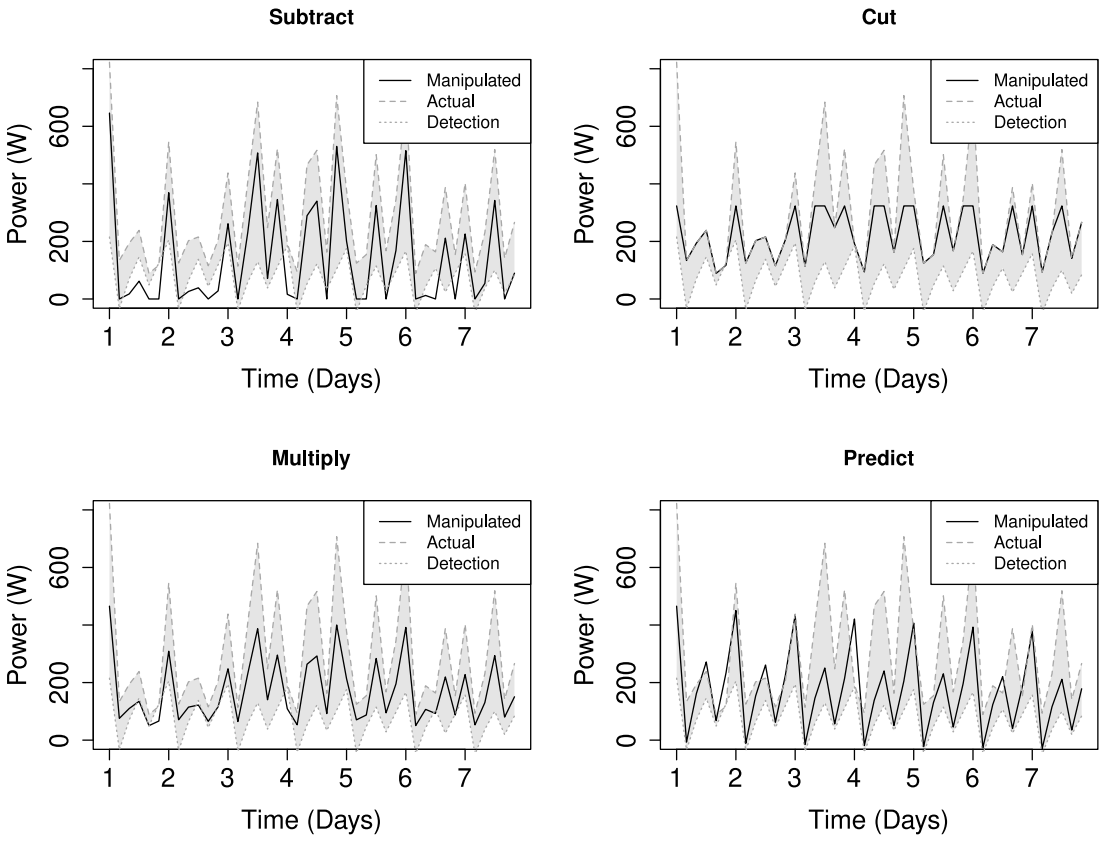


Figure 7.7: Four different manipulation methods.

- Subtract: Subtracts a constant in range [0, max(Power)] from every measurement and set negative values to zero.
- Cut: Sets any measurement above a constant in range [0, max(Power)] to the value of this constant.
- Multiply: Multiplies a constant in range [0,1] with every measurement.
- Predict: Replaces the load curve with the prediction.

In case of 'Predict', the Holt Winters algorithm is used with 45 instead of 48 training values, to simulate that the adversary does not know the exact parameters of anomaly detection system. Furthermore, a constant in range [0, max(Power)] is subtracted from the prediction.

Figure 7.8 shows the number of anomalous measurements created by each manipulation method (avg. 100 experiments). The y-axis shows the amount of measurements below the threshold and the x-axis shows the energy theft in percent. Hence, 100% on the y-axis means that every single measurements of the load curve is anomalous. The method 'cut', where energy peaks above a certain power are cut off, works well because the anomaly detection method only considered a lower threshold. Some damage could be avoided by adding an upper threshold for a minimum energy demand. However, in practise such a threshold may increase the

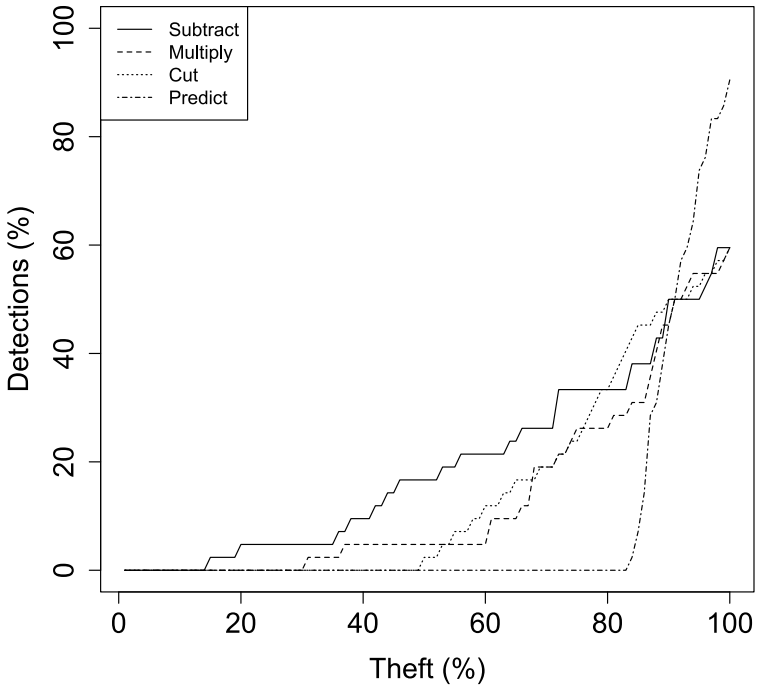


Figure 7.8: Energy theft vs measurements lower than confidence band.

number of false positives. It is not surprising that the adversary can gain the most benefit by predicting the energy demand and mimicking the statistics of the anomaly detection system.

Next, these four methods are compared for the Naive Bayes anomaly detection. However, to simplify the experiment and comparison to the previous anomaly detection, it is assumed that a manipulated measurement is anomalous if assigned to a different cluster centre. Hence, the adversary aims to decrease the power without changing the input values of Naive Bayes algorithm.

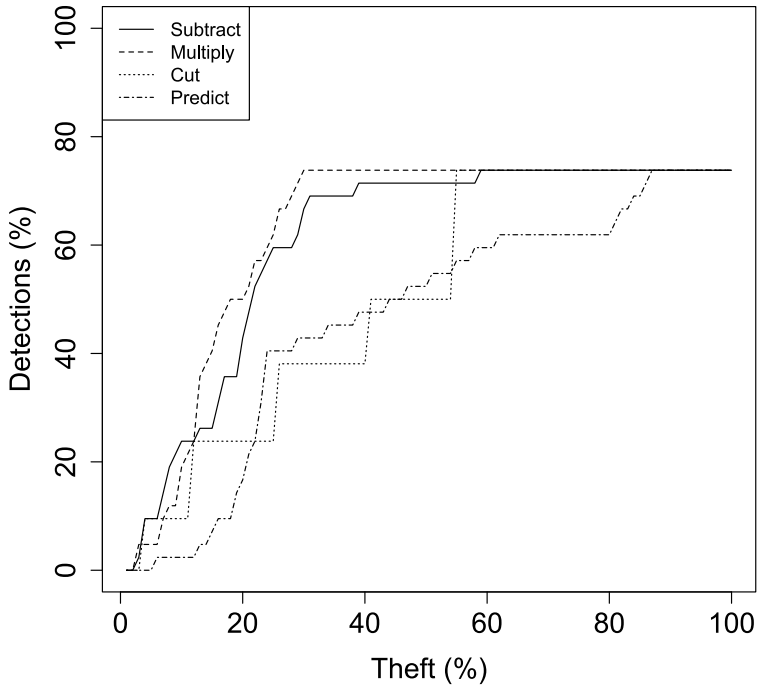


Figure 7.9: Energy theft vs amount of changed clusters.

For the method 'predict', the adversary applies the k-means algorithm on the previous day to find the cluster centres. Each measurement can be set to the minimum value of the cluster, or if all measurements are already at the minimum the highest power value was set to zero. Figure 7.9 shows the output of each method, again the y-axis shows the amount of measurements below the threshold and the x-axis shows the energy theft in percent. Here, the values on the y-axis only reached approximately 75%, because the measurements which already belonged to the cluster with smallest power value were not changed by the adversary.

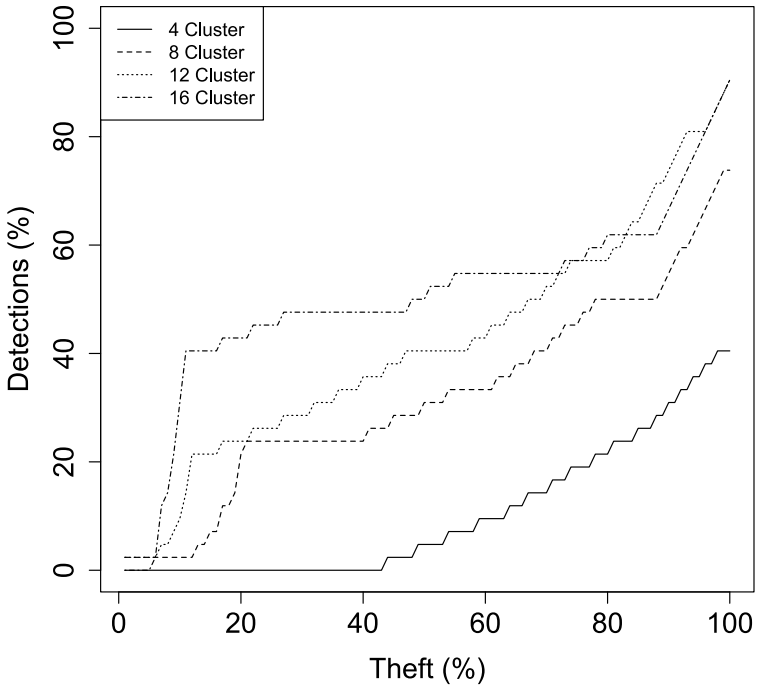


Figure 7.10: Energy theft vs amount of changed clusters.

The method seems much more resilient against manipulation than the previous method, because in contrast to the Holt Winters prediction smaller changes are detected. Note that, these results do not necessarily mean that the anomaly detection is better in general as this sensitivity also makes it more difficult to fine-tune the system to the actual behaviour of a household – a comparison of the anomaly detection systems can be found in the previous chapter. Here, the performance of the 'predict' method is not clearly better than the other methods. Figure 7.10 shows that the performance depends upon the number of clusters

(or number of appliances in case of a different NIALM method). The figure shows manipulated measurements that can be detected (y-axis) vs amount of energy theft (x-axis) with different amounts of clusters for the method 'predict' (line type). The result is reasonable, because the more clusters there are, the smaller the difference to the next cluster, which can be subtracted in the proposed stealthy theft method.

7.4 Discussion

To investigate RQ6, this chapter showcased stealthy energy theft, by means of mimicking the characteristics expected by the anomaly detection system, with two different anomaly detection approaches. It can be concluded, that stealthy energy theft is under certain conditions not detectable, e.g. if the detection algorithm is known and the adversary can arbitrarily manipulate the electricity. However, it is possible to limit the amount of undetectable energy theft. Here, the threshold of undetectable theft was called the error margin of the anomaly detection system, which was in case of the Holt Winters method the prediction error estimated by the RMSE and in case of the Naive Bayes method the range of each cluster.

Chapter 8

Conclusion

The objective of this thesis was to design and evaluate an anomaly detection system suitable for the smart grid. This work concluded that anomaly detection can help to detect the manipulation of energy demand or the concrete scenario of energy theft early on, which is one of the current research topics related to the security of the smart grid. The following presents the conclusions and points out contributions and research questions as well as some suggestions and directions for future work.

8.1 Achievements

This work analysed available data and focused on the characteristics of power load curves, which was used to introduce several metrics to extract the human activity and compared methods to monitor the metrics. In

order to fully utilize the entropy inspired metric, an anomaly detection method based on Holt Winters prediction was introduced and evaluated in comparison to other state of the art methods based on Naive Bayes and XMR. One of the major contributions is the comprehensive analysis of the different 'dimensions' that can be used to model the normal behaviour. It was found that, with feasible metrics, it is possible to compare the consumption of similar smart meters to detect anomalies. Last but not least, the thesis showcased stealthy energy theft as one of the limitations of anomaly detection and pointed out how to mitigate such scenarios.

The early chapters showcased that any action that jeopardizes the confidentiality, integrity or availability of the smart grid are threats. While many threat scenarios are not different from the security of other IoT devices, this work briefly introduced the architecture with interfaces to different networks as well as the possibility to manipulate measurements, which is unique to smart meters and the smart grid. The thesis provided a brief threat taxonomy and motivated that encryption and physical security are not sufficient to ensure safety and security of smart meters. Energy theft was introduced as one of the threats in the smart grid and classified in intrusive and non-intrusive tampering methods.

The next chapter provided an extensive analysis of energy demand,

including measurements, available datasets, simulations and appliance load curves with a comprehensive literature review. Two characteristics were analysed in detail, namely the human activity, which defines the on/off time of appliances, and individual appliance load curves, which define the power when all active appliances are summarized. Both characteristics together are often used to simulate energy demand and can also be extracted from an aggregated load curve. The usage of the human activity was confirmed by evaluating characteristic time periods of load curves using the entropy as a metric. The reasoning here was, that characteristic energy demand shapes are often gathered in few periods per day and repeat daily according to the consumers habits. The results of the experiments showed that periods without consumption can often be identified during the same time spans over several days and hence indicate characteristic human activity.

Next, the feasibility of an entropy inspired metric to extract human activity was evaluated. The thesis introduced two methods which reflect changes in energy demand, namely the sliding window entropy and the interval entropy. The experiments utilizing the ECO dataset presented an analysis of the corresponding parameters and highlighted the accuracy of this approach in comparison to other statistical methods. As a non-intrusive approach without a-priori data, the method archived re-

sults which are overall better than comparable methods with a moderate number of input values, which is suitable for real world applications.

Based on the previous metrics, the thesis continued with a description of the challenges and prerequisites of comparing multiple households and proposed some characteristics that are especially suitable to compare multiple households. The chapter demonstrated three exemplary features, derived from raw energy demand, which are normalized to a fixed range over a time window, so that different load curves can be compared. The experiments examined the statistical influence of parameters, presented a systematic approach to fine-tune and adjust them, and evaluated the quality of each features to detect energy theft.

The next part showed that multiple data sources can indeed unveil otherwise hidden outliers. For the two scenarios of energy theft, the results showed detection rates above 90%. The chapter showcased the advantage of using several data sources and designed a method to remove the daily pattern from multiple sources while preserving outliers which represent energy theft. The experiments compared the proposed metric with an alternative distance-based metric and showed that the entropy-inspired metric is especially robust in presence of multiple outliers. Furthermore, the experiments evaluated the performance of the entropy-inspired method against two other state of the art anomaly de-

tection methods.

Last but not least, the thesis showcased stealthy energy theft methods, which can mimic the expected characteristics of energy demand to avoid detection. To point out the constraints and inherent limitations of anomaly detection with regard to sophisticated and stealthy energy theft methods, the chapter described the conditions, such as the digital access to corrupt smart meters and knowledge of the anomaly detection method, to execute such scenarios. The thesis introduced a concept to mimic the expected behaviour for two exemplary anomaly detection models and compared the maximum amount of stolen energy under different conditions.

8.2 Future Work

There are many interesting topics, related to security and anomaly detection in smart grids, aside from the questions answered in this study. Unfortunately, the access to real world data is often difficult due to privacy reasons. It would be interesting to see public data sets with the corresponding weather data or socio-economic data and also data sets which include energy production as today's energy production shifts more and more towards the low-voltage network. In general, data which is not available from end consumer smart meters falls outside this study,

but the author is well aware that data from power plants, transmission stations and distribution stations can extend and improve anomaly detection. While it would be interesting to examine, the usage of external data such as weather data or socio-economic data to improve the prediction quality was not discussed in detail. As future prospects, the author supposes to evaluate the usage of additional measurements, which are correlated to the power. Throughout the work, a focus was on the design of metrics, derived from raw data. However, automatic parameter tuning to optimize the detection with a certain set of households was neglected and could further improve the detection rate. Furthermore, clustering load curves together could further improve the results of anomaly detection. Anomalies in the network communication between smart meters were excluded from this study. But the homogenous network traffic in the smart grid can provide opportunities for a better detection quality in comparison to the regular internet with innumerable applications and protocols. A limitation specific to the anomaly detection method presented in this work is the low output resolution, because the method needs to summarise several measurements within a time window to compute the metric. Lowering this requirement would be an interesting future work. While this work introduced the concept to reduce energy theft with sophisticated and stealthy tampering

methods, it did not focus on methods to harden the anomaly detection system against such methods or attempts to compromise the anomaly detection itself, which are well-known from other areas. An in-depth analysis of possible architectures and deployment scenarios of anomaly detection sensors would extend this work very well. Multi-sensor setups or peer-to-peer structures could contribute to anomaly detection in the smart grid. This work also excludes byzantine attacks and any scenarios which only work by gaining control of several compromised smart meters. However, anomaly detection systems similar to the methods introduced in this work can also be adapted to these scenarios.

Terms and Abbreviations

AMPd2 Almanac of Minutly Power. 59

ANN Artificial Neural Network. 37–39, 102, 103

ARIMA Autoregressive Integrated Moving Average. 36, 37

AUC Area Under the Curve. 52, 53, 95–97, 136, 140–142, 146, 147, 149, 158–160, 163

CDA Conditional Demand Analysis. 36, 37, 102, 103

D2M Distance to Maximum. 140, 142, 151

DSM Demand Side Management. 16, 17, 36

ECO Electricity Consumption & Occupancy. 57, 59–63, 81, 84, 93, 98, 109, 110, 126, 148, 155, 157

FN False Negative. 50, 122

FP False Positive. 49, 50, 122

GA Genetic Algorithm. 37, 38

HAN Home Area Network. 31, 41

HMM Hidden Markov Model. 102, 104

iAWE Indian Dataset for Ambient Water and Energy. 59, 60

IHEPC Individual Household Electric Power Consumption. 59

KNN K-Nearest Neighbours. 104

LDA Linear Discriminant Analysis. 104

LMN Local Metrological Network. 31

NIALM Non-Intrusive Appliance Load Monitoring. 15, 17, 32, 47, 57, 60, 101, 102, 144, 145, 162

NIST National Institute of Standards and Technology. 29, 40

PCA Principal Component Analysis. 35, 104

REDD Reference Energy Disaggregation Data. 60

RMSE Root-Mean-Square Error. 137–139, 143, 158, 159, 161

ROC Receiver Operating Characteristic. 51, 53, 95, 136, 141, 142, 159

sd Standard Deviation. 74, 119, 122, 133, 134, 137, 150

SMGW Smart Meter Gateway. 30, 31, 40, 41

SVM Support Vector Machine. 36–38, 102, 104

TLC Typical Load Classification. 15, 33, 57, 101, 105

TN True Negative. 50, 122

TP True Positive. 50, 122

WAN Wide Area Network. 31

References

Aigner, D. J., Sorooshian, C., & Kerwin, P. (1984). Conditional demand analysis for estimating residential end-use load profiles. *The Energy Journal*, *5*(3), 81–97.

Anderson, B., Lin, S., Newing, A., Bahaj, A., & James, P. (2017). Electricity consumption and household characteristics: Implications for census-taking in a smart metered future. *Computers, Environment and Urban Systems*, *63*, 58–67.

Andrade, E. O., Sampaio, I. G., Viterbo, J., Silva, J. M., & Boscarioli, C. (2016). Profiling household consumption with clustering algorithms. In *Proceedings of the 7th brazilian symposium on information systems (sbsi)* (p. 7).

Andrysiak, T., Saganowski, Ł., & Kiedrowski, P. (2017). Anomaly detection in smart metering infrastructure with the use of time series analysis. *Journal of Sensors*, *2017*.

Anu, J., Agrawal, R., Seay, C., & Bhattacharya, S. (2015). Smart grid security risks. In *Proceedings of the 12th international conference on information technology-new generations (itng)* (pp. 485–489).

Armstrong, M. M., Swinton, M. C., Ribberink, H., Beausoleil Morrison, I., & Millette, J. (2009). Synthetically derived profiles for representing occupant-driven electric loads in canadian housing. *Journal of Building Performance Simulation*, *2*(1), 15–30.

Arroyo, J., San Roque, A. M., Maté, C., & Sarabia, A. (2007). Exponential smoothing methods for interval time series. In *Proceedings of the 1st european symposium on time series prediction* (pp. 231–240).

Arshad, N., Ali, U., & Javed, F. (2013). A highly configurable simulator for assessing energy usage. *Energy Procedia*, *42*, 308–317.

Ashrafuzzaman, M., Chakhchoukh, Y., Jillepalli, A. A., Tosic, P. T., de Leon, D. C., Sheldon, F. T., & Johnson, B. K. (2018). Detecting stealthy false data injection attacks in power

grids using deep learning. In *Proceedings of the 14th international wireless communications & mobile computing conference (iwcmc)* (pp. 219–225).

Aydinalp, M., Ugursal, V. I., & Fung, A. S. (2002). Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks. *Applied energy*, *71*(2), 87–110.

Bandim, C., Alves, J., Pinto, A., Souza, F., Loureiro, M., Magalhaes, C., & Galvez-Durand, F. (2003). Identification of energy theft and tampered meters using a central observer meter: a mathematical approach. In *Proceedings of the 1th conference on transmission and distribution (t&d)* (Vol. 1, pp. 163–168).

Banos, R., Manzano Agugliaro, F., Montoya, F., Gil, C., Alcayde, A., & Gómez, J. (2011). Optimization methods applied to renewable and sustainable energy: a review. *Renewable and Sustainable Energy Reviews*, *15*(4), 1753–1766.

Baranski, M., & Voss, J. (2004). Genetic algorithm for pattern detection in nialm systems. In *Proceedings of the 3rd international conference on systems, man and cybernetics (smc)* (Vol. 4, pp. 3462–3468).

Barker, S., Mishra, A., Irwin, D., Cecchet, E., Shenoy, P., & Albrecht, J. (2012). Smart*: An open data set and tools for enabling research in sustainable homes. In *Proceedings of the 18th workshop on data mining applications in sustainability (sigkdd)* (p. 112).

Batra, N., Gulati, M., Singh, A., & Srivastava, M. B. (2013). It's different: insights into home energy consumption in india. In *Proceedings of the 5th workshop on embedded systems for energy-efficient buildings (buildsys)* (pp. 1–8).

Beckel, C., Kleiminger, W., Cicchetti, R., Staake, T., & Santini, S. (2014). The eco data set and the performance of non-intrusive load monitoring algorithms. In *Proceedings of the 6th conference on embedded systems for energy-efficient buildings (buildsys)* (pp. 80–89).

Beckel, C., Sadamori, L., & Santini, S. (2013). Automatic socio-economic classification of households using electricity consumption data. In *Proceedings of the 4th international conference on future energy systems (e-energy)* (pp. 75–86).

Bettencourt, L. M., Hagberg, A. A., & Larkey, L. B. (2007). Separating the wheat from the chaff: practical anomaly detection schemes in ecological applications of distributed sensor networks. In *Distributed computing in sensor systems* (pp. 223–239). Springer.

Bhattacharjee, S., & Das, S. K. (2018). Detection and forensics against stealthy data falsification

in smart metering infrastructure. *IEEE Transactions on Dependable and Secure Computing*.

Bobba, R. B., Rogers, K. M., Wang, Q., Khurana, H., Nahrstedt, K., & Overbye, T. J. (2010). Detecting false data injection attacks on dc state estimation. In *Proceedings of the 1st workshop on secure control systems (scs)* (Vol. 1, pp. 1–9).

Bohi, D. R., & Zimmerman, M. B. (1984). An update on econometric studies of energy demand behavior. *Annual Review of Energy*, *9*(1), 105–154.

Bouché, J., Hock, D., & Kappes, M. (2016). On the performance of anomaly detection systems uncovering traffic mimicking covert channels. In *Proceedings of the 11th international network conference (inc)* (pp. 19–24).

Braun, H., Buddha, S. T., Krishnan, V., Spanias, A., Tepedelenlioglu, C., Yeider, T., & Takehara, T. (2012). Signal processing for fault detection in photovoltaic arrays. In *Proceedings of the 37th international conference on acoustics, speech and signal processing (icassp)* (pp. 1681–1684).

Capasso, A., Grattieri, W., Lamedica, R., & Prudenzi, A. (1994). A bottom-up approach to residential load modeling. *IEEE Transactions on Power Systems*, *9*(2), 957–964.

Cárdenas, A. A., Amin, S., Schwartz, G., Dong, R., & Sastry, S. (2012). A game theory model for electricity theft detection and privacy-aware control in ami systems. In *Proceedings of the 50th conference on communication, control, and computing (allerton)* (pp. 1830–1837).

Carroll, J., Lyons, S., & Denny, E. (2014). Reducing household electricity demand through smart metering: the role of improved information about energy saving. *Energy Economics*, *45*, 234–243.

Casenove, M. (2015). Exfiltrations using polymorphic blending techniques: analysis and countermeasures. In *Proceedings of the 7th international conference on cyber conflict (cycon)* (pp. 217–230).

Chang, J., Leung, D. Y., Wu, C., & Yuan, Z. (2003). A review on the energy production, consumption, and prospect of renewable energy in china. *Renewable and Sustainable Energy Reviews*, *7*(5), 453–468.

Chen, C., & Liu, L. M. (1993). Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, *88*(421), 284–297.

Chen, D., Barker, S., Subbaswamy, A., Irwin, D., & Shenoy, P. (2013). Non-intrusive occupancy monitoring using smart meters. In *Proceedings of the 5th workshop on embedded systems for*

energy-efficient buildings (buildsys) (pp. 1–8).

Chicco, G., & Ilie, I. S. (2009). Support vector clustering of electrical load pattern data. *IEEE Transactions on Power Systems*, *24*(3), 1619–1628.

Chicco, G., Ionel, O. M., & Porumb, R. (2013). Electrical load pattern grouping based on centroid model with ant colony clustering. *IEEE Transactions on Power Systems*, *28*(2), 1706–1715.

Chicco, G., Napoli, R., & Piglione, F. (2006). Comparisons among clustering techniques for electricity customer classification. *IEEE Transactions on Power Systems*, *21*(2), 933–940.

Cho, M., Hwang, J., & Chen, C. (1995). Customer short term load forecasting by using arima transfer function model. In *Proceedings of the 1st international conference on energy management and power delivery (empd)* (Vol. 1, pp. 317–322).

Connolly, D., Lund, H., Mathiesen, B. V., & Leahy, M. (2010). A review of computer tools for analysing the integration of renewable energy into various energy systems. *Applied Energy*, *87*(4), 1059–1082.

Corona, I., Giacinto, G., & Roli, F. (2013). Adversarial attacks against intrusion detection systems: taxonomy, solutions and open issues. *Information Sciences*, *239*, 201–225.

Dabrowski, A., Ullrich, J., & Weippl, E. R. (2017). Grid shock: coordinated load-changing attacks on power grids: the non-smart power grid is vulnerable to cyber attacks as well. In *Proceedings of the 33rd annual computer security applications conference (acsac)* (pp. 303–314).

Datta, D., Tassou, S., & Marriott, D. (2000). Application of neural networks for the prediction of the energy consumption in a supermarket. In *Proceedings of the 1st international conference on clima (clima)* (pp. 98–107).

Delgado-Gomes, V., Martins, J. F., Lima, C., & Borza, P. N. (2015). Smart grid security issues. In *Proceedings of the 9th international conference on compatibility and power electronics (cpe)* (pp. 534–538).

Denning, D. E. (1987). An intrusion-detection model. *IEEE Transactions on Software Engineering*, *13*(2), 222–232.

Depuru, S. S. S. R., Wang, L., & Devabhaktuni, V. (2011). Support vector machine based data classification for detection of electricity theft. In *Proceedings of the 1th conference on power systems (psce)* (pp. 1–8).

Dienst, S., Schmidt, J., & Kühne, S. (2013). Case study: condition assessment of a photovoltaic power plant using change-point analysis. In *Proceedings of the 2nd international conference*

on smart grids and green it systems (smartgreens) (pp. 159–164).

Druckman, A., & Jackson, T. (2008). Household energy consumption in the uk: a highly geographically and socio-economically disaggregated model. *Energy Policy*, *36*(8), 3177–3192.

Dutta, G., & Mitra, K. (2017). A literature review on dynamic pricing of electricity. *Journal of the Operational Research Society*, *68*(10), 1131–1145.

Ericsson, G. N. (2010). Cyber security and power system communication–essential parts of a smart grid infrastructure. *IEEE Transactions on Power Delivery*, *25*(3), 1501–1507.

Esmalifalak, M., Nguyen, H., Zheng, R., & Han, Z. (2011). Stealth false data injection using independent component analysis in smart grid. In *Proceedings of the 2nd international conference on smart grid communications (smartgridcomm)* (pp. 244–248).

Fengming, Z., Shufang, L., Zhimin, G., Bo, W., Shiming, T., & Mingming, P. (2017). Anomaly detection in smart grid based on encoder-decoder framework with recurrent neural network. *The Journal of China Universities of Posts and Telecommunications*, *24*(6), 67–73.

Foucquier, A., Robert, S., Suard, F., Stéphan, L., & Jay, A. (2013). State of the art in building modelling and energy performances prediction: a review. *Renewable and Sustainable Energy Reviews*, *23*, 272–288.

Freedman, D., Pisani, R., & Purves, R. (2007). *Statistics 4th edition.* Norton & Company.

Frolik, J., Abdelrahman, M., & Kandasamy, P. (2001). A confidence-based approach to the self-validation, fusion and reconstruction of quasi-redundant sensor data. *IEEE Transactions on Instrumentation and Measurement*, *50*(6), 1761–1769.

Ghanbari, A., Kazemi, S. M., Mehmanpazir, F., & Nakhostin, M. M. (2013). A cooperative ant colony optimization-genetic algorithm approach for construction of energy demand forecasting knowledge-based expert systems. *Knowledge-Based Systems*, *39*, 194–206.

Giani, A., Bitar, E., Garcia, M., McQueen, M., Khargonekar, P., & Poolla, K. (2011). Smart grid data integrity attacks: characterizations and countermeasures π. In *Proceedings of the 2nd international conference on smart grid communications (smartgridcomm)* (pp. 232–237).

Grandjean, A., Adnot, J., & Binet, G. (2012). A review and an analysis of the residential electric load curve models. *Renewable and Sustainable Energy Reviews*, *16*(9), 6539–6565.

Gungor, V. C., Sahin, D., Kocak, T., Ergüt, S., Buccella, C., Cecati, C., & Hancke, G. P. (2012). Smart grid and smart homes: key players and pilot projects. *IEEE Industrial Electronics Magazine*, *6*(4), 18–34.

Hart, G. W. (1992). Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, *80*(12), 1870–1891.

Hébrail, G., & Bérard, A. (2012). Individual household electric power consumption data set. *É. d. France, Ed., ed: UCI Machine Learning Repository*.

Hirst, E., Lin, W., & Cope, J. (1977). Residential energy use model sensitive to demographic, economic, and technological factors. *Quarterly Review of Economics & Business*, *17*(2).

Hock, D., Kappes, M., & Ghita, B. (2020). Using multiple data sources to detect manipulated electricity meter by an entropy-inspired metric. *Sustainable Energy, Grids and Networks*, *21*, 100290.

Hong, W. C. (2009). Electric load forecasting by support vector model. *Applied Mathematical Modelling*, *33*(5), 2444–2454.

Hunt, L. C., Judge, G., & Ninomiya, Y. (2003). Underlying trends and seasonality in uk energy demand: a sectoral analysis. *Energy Economics*, *25*(1), 93–118.

Illera, A., & Vidal, J. (2014). Lights off! the darkness of the smart meters. *BlackHat Europe*.

Jain, P., & Tripathi, P. (2013). Scada security: a review and enhancement for dnp3 based systems. *CSI Transactions on ICT*, *1*(4), 301–308.

Jokar, P., Arianpoo, N., & Leung, V. C. (2015). Electricity theft detection in ami using customers' consumption patterns. *IEEE Transactions on Smart Grid*, *7*(1), 216–226.

Jota, P. R., Silva, V. R., & Jota, F. G. (2011). Building load management using cluster and statistical analyses. *Electrical Power & Energy Systems*, *33*(8), 1498–1505.

Kalogirou, S. A., & Bojic, M. (2000). Artificial neural networks for the prediction of the energy consumption of a passive solar building. *Energy*, *25*(5), 479–491.

Kavaklioglu, K., Ceylan, H., Ozturk, H. K., & Canyurt, O. E. (2009). Modeling and prediction of turkey's electricity consumption using artificial neural networks. *Energy Conversion and Management*, *50*(11), 2719–2727.

Kelly, J., & Knottenbelt, W. (2015). The uk-dale dataset, domestic appliance-level electricity demand and whole-house demand from five uk homes. *Scientific data*, *2*, 150007.

Kendall, M., & Stuart, A. (1983). *The advanced theory of statistics* (Vol. 3). Griffin.

Kim, H., Marwah, M., Arlitt, M., Lyon, G., & Han, J. (2011). Unsupervised disaggregation of low frequency power measurements. In *Proceedings of the 11th international conference on data mining (icdmw)* (pp. 747–758).

Kleiminger, W., Beckel, C., & Santini, S. (2015a). Household occupancy monitoring using electricity meters. In *Proceedings of the 3rd international joint conference on pervasive and ubiquitous computing (ubicomp).* Osaka, Japan.

Kleiminger, W., Beckel, C., & Santini, S. (2015b). Household occupancy monitoring using electricity meters. In *Proceedings of the 3rd international joint conference on pervasive and ubiquitous computing (ubicomp)* (pp. 975–986).

Kolter, J. Z., & Ferreira Jr., J. (2011). A large-scale study on predicting and contextualizing building energy usage. In *Proceedings of the 25th conference on artificial intelligence (aaai).*

Kolter, J. Z., & Johnson, M. J. (2011). Redd: A public data set for energy disaggregation research. In *Proceedings of the 17th workshop on data mining applications in sustainability (sigkdd)* (Vol. 25, pp. 59–62).

Kosut, O., Jia, L., Thomas, R. J., & Tong, L. (2010). Limiting false data attacks on power system state estimation. In *Proceedings of the 44th conference on information sciences and systems (ciss)* (pp. 1–6).

Labandeira, X., Labeaga Azcona, J. M., & Rodríguez Méndez, M. (2006). A residential energy demand system for spain. *The Energy Journal*, *27*(2).

Lamport, L., Shostak, R., & Pease, M. (1982). The byzantine generals problem. *Transactions on Programming Languages and Systems*, *4*(3), 382–401.

Liu, D., Niu, D., Wang, H., & Fan, L. (2014). Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. *Renewable Energy*, *62*, 592–597.

Liu, J., Xiao, Y., Li, S., Liang, W., & Chen, C. (2012). Cyber security and privacy issues in smart grids. *IEEE Communications Surveys & Tutorials*, *14*(4), 981–997.

Liu, Y., Liu, T., Sun, H., Zhang, K., & Liu, P. (2020). Hidden electricity theft by exploiting multiple-pricing scheme in smart grids. *IEEE Transactions on Information Forensics and Security*, *15*, 2453–2468.

Liu, Y., Ning, P., & Reiter, M. K. (2011). False data injection attacks against state estimation in electric power grids. *Transactions on Information and System Security*, *14*(1), 13.

Livani, H., & Evrenosoglu, C. Y. (2013). A machine learning and wavelet-based fault location method for hybrid transmission lines. *IEEE Transactions on Smart Grid*, *5*(1), 51–59.

Makonin, S. (2016). Investigating the switch continuity principle assumed in non-intrusive load

monitoring (nilm). In *Proceedings of the 29th canadian conference on electrical and computer engineering (ccece)* (pp. 1–4).

Makonin, S., Ellert, B., Bajić, I. V., & Popowich, F. (2016). Electricity, water, and natural gas consumption of a residential house in canada from 2012 to 2014. *Scientific data*, *3*.

Mashima, D., & Cárdenas, A. A. (2012). Evaluating electricity theft detectors in smart grid networks. In *Proceedings of the 12th international workshop on recent advances in intrusion detection (raid)* (pp. 210–229).

McDaniel, P., & McLaughlin, S. (2009). Security and privacy challenges in the smart grid. *IEEE Security & Privacy*, *7*(3), 75–77.

McLaughlin, S., Holbert, B., Fawaz, A., Berthier, R., & Zonouz, S. (2013). A multi-sensor energy theft detection framework for advanced metering infrastructures. *IEEE Journal on Selected Areas in Communications*, *31*(7), 1319–1330.

McLaughlin, S., Holbert, B., Zonouz, S., & Berthier, R. (2012). Amids: A multi-sensor energy theft detection framework for advanced metering infrastructures. In *Proceedings of the 3rd international conference on smart grid communications (smartgridcomm)* (pp. 354–359).

McLoughlin, F. (2013). *Characterising domestic electricity demand for customer load profile segmentation* (Unpublished doctoral dissertation). Dublin Institute of Technology.

McLoughlin, F., Duffy, A., & Conlon, M. (2013). Evaluation of time series techniques to characterise domestic electricity demand. *Energy*, *50*, 120–130.

Metke, A. R., & Ekl, R. L. (2010). Smart grid security technology. In *Proceedings of the 19th conference on innovative smart grid technologies (isgt)* (pp. 1–7).

Mihalakakou, G., Santamouris, M., & Tsangrassoulis, A. (2002). On the energy consumption in residential buildings. *Energy and Buildings*, *34*(7), 727–736.

Molina-Markham, A., Shenoy, P., Fu, K., Cecchet, E., & Irwin, D. (2010). Private memoirs of a smart meter. In *Proceedings of the 2nd workshop on embedded sensing systems for energy-efficiency in building (buildsys)* (pp. 61–66).

Monacchi, A., Egarter, D., Elmenreich, W., D'Alessandro, S., & Tonello, A. M. (2014). Greend: an energy consumption dataset of households in italy and austria. In *Proceedings of the 5th international conference on smart grid communications (smartgridcomm)* (pp. 511–516).

Mookiah, L., Dean, C., & Eberle, W. (2017). Graph-based anomaly detection on smart grid data. In *Proceedings of the 30th international flairs conference (flairs).*

Murray, D., Stankovic, L., & Stankovic, V. (2017). An electrical load measurements dataset of united kingdom households from a two-year longitudinal study. *Scientific data*, *4*, 160122.

Nagi, J., Yap, K., Tiong, S., Ahmed, S., & Mohammad, A. (2008). Detection of abnormalities and electricity theft using genetic support vector machines. In *Proceedings of the 6th conference on technology, knowledge, and society (tencon)* (pp. 1–6).

Newing, A., Anderson, B., Bahaj, A., & James, P. (2016). The role of digital trace data in supporting the collection of population statistics–the case for smart metered electricity consumption data. *Population, Space and Place*, *22*(8), 849–863.

Nguyen, T. A., & Aiello, M. (2013). Energy intelligent buildings based on user activity: a survey. *Energy and buildings*, *56*, 244–257.

Nizar, A., Dong, Z., & Wang, Y. (2008). Power utility nontechnical loss analysis with extreme learning machine method. *IEEE Transactions on Power Systems*, *23*(3), 946–955.

Ozturk, H. K., Canyurt, O. E., Hepbasli, A., & Utlu, Z. (2004). Residential-commercial energy input estimation based on genetic algorithm (ga) approaches: an application of turkey. *Energy and Buildings*, *36*(2), 175–183.

Paatero, J. V., & Lund, P. D. (2006). A model for generating household electricity load profiles. *International Journal of Energy Research*, *30*(5), 273–290.

Parti, M., & Parti, C. (1980). The total and appliance-specific conditional demand for electricity in the household sector. *The Bell Journal of Economics*, 309–321.

Pasqualetti, F., Carli, R., & Bullo, F. (2011). A distributed method for state estimation and false data detection in power networks. In *Proceedings of the 2nd international conference on smart grid communications (smartgridcomm)* (pp. 469–474).

Peterson, W., Birdsall, T., & Fox, W. (1954). The theory of signal detectability. *Transactions of the IRE Professional Group on Information Theory*, *4*(4), 171-212. doi: 10.1109/TIT.1954.1057460

Price, P. (2010). Methods for analyzing electric load shape and its variability. *Lawrence Berkeley National Laboratory*.

Raciti, M., & Nadjm Tehrani, S. (2013). Embedded cyber-physical anomaly detection in smart meters. In *Critical information infrastructures security* (pp. 34–45). Springer.

Reinhardt, A., Baumann, P., Burgstahler, D., Hollick, M., Chonov, H., Werner, M., & Steinmetz, R. (2012). On the accuracy of appliance identification based on distributed load metering

data. In *Proceedings of the 2nd conference on sustainable internet and ict for sustainability (sustainit)* (pp. 1–9).

Richardson, I., Thomson, M., Infield, D., & Clifford, C. (2010). Domestic electricity use: a high-resolution energy demand model. *Energy and Buildings*, *42*(10), 1878–1887.

Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, *278*(6), 2039–2049.

Ross, R. S., McEvilley, M., & Oren, J. C. (2018). *Systems security engineering: considerations for a multidisciplinary approach in the engineering of trustworthy secure systems* (Tech. Rep.). 100 Bureau Drive, Gaithersburg, MD 20899: NIST.

Rossi, B., Chren, S., Buhnova, B., & Pitner, T. (2016). Anomaly detection in smart grid data: an experience report. In *Proceedings of the 15th international conference on systems, man and cybernetics (smc)* (pp. 002313–002318).

Ruzzelli, A. G., Nicolas, C., Schoofs, A., & O'Hare, G. M. (2010). Real-time recognition and profiling of appliances through a single electricity sensor. In *Proceedings of the 7th conference on sensor mesh and ad hoc communications and networks (secon)* (pp. 1–9).

Sadeghi, H., Zolfaghari, M., & Heydarizade, M. (2011). Estimation of electricity demand in residential sector using genetic algorithm approach. *International Journal of Industrial Engineering*, *22*(1).

Saitoh, T., Osaki, T., Konishi, R., & Sugahara, K. (2010). Current sensor based home appliance and state of appliance recognition. *Journal of Control, Measurement, and System Integration*, *3*(2), 86–93.

Salinas, S., Li, M., & Li, P. (2013). Privacy-preserving energy theft detection in smart grids: a p2p computing approach. *IEEE Journal on Selected Areas in Communications*, *31*(9), 257–267.

Shah, A. J., Desrosiers, C., & Sabourin, R. (2015). Contextual anomaly detection using log-linear tensor factorization. In *Advances in knowledge discovery and data mining* (pp. 165–176). Springer.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423.

Spirić, J. V., Dočić, M. B., & Stanković, S. S. (2015). Fraud detection in registered electricity time series. *International Journal of Electrical Power & Energy Systems*, *71*, 42–50.

Stokes, M. (2005). *Removing barriers to embedded generation: a fine-grained load model to support low voltage network performance analysis* (Unpublished doctoral dissertation). De Montfort University.

Suganthi, L., & Samuel, A. A. (2012). Energy models for demand forecasting–a review. *Renewable and Sustainable Energy Reviews*, *16*(2), 1223–1240.

Swan, L. G., & Ugursal, V. I. (2009). Modeling of end-use energy consumption in the residential sector: a review of modeling techniques. *Renewable and Sustainable Energy Reviews*, *13*(8), 1819–1835.

Szmit, M., Szmit, A., Adamus, S., & Bugała, S. (2012). Usage of holt-winters model and multilayer perceptron in network traffic modelling and anomaly detection. *Informatica*, *36*(4).

Tiedemann, K. (2007). Using conditional demand analysis to estimate residential energy use and energy savings. In *Proceedings of the 1st conference on saving energy (eceee).*

Train, K., Herriges, J., & Windle, R. (1985). Statistically adjusted engineering (sae) models of end-use load curves. *Energy*, *10*(10), 1103–1111.

Trierweiler Ribeiro, G., Guilherme Sauer, J., Fraccanabbia, N., Cocco Mariani, V., & dos Santos Coelho, L. (2020). Bayesian optimized echo state network applied to short-term load forecasting. *Energies*, *13*(9), 2390.

Tzafestas, S., & Tzafestas, E. (2001). Computational intelligence techniques for short-term electric load forecasting. *Journal of Intelligent and Robotic Systems*, *31*(1-3), 7–68.

Urbina, D. I., Giraldo, J. A., Cardenas, A. A., Tippenhauer, N. O., Valente, J., Faisal, M., ... Sandberg, H. (2016). Limiting the impact of stealthy attacks on industrial control systems. In *Proceedings of the 23rd conference on computer and communications security (ccs)* (pp. 1092–1105).

Von Dollen, D. (2009). *Report to nist on the smart grid interoperability standards roadmap* (Tech. Rep.). U.S. Department of Energy, Office of Electricity, Washington DC: Electric Power Research Institute (EPRI) and National Institute of Standards and Technology.

Vranken, I., Baudry, J., Aubinet, M., Visser, M., & Bogaert, J. (2015). A review on the use of entropy in landscape ecology: heterogeneity, unpredictability, scale dependence and their links with thermodynamics. *Landscape ecology*, *30*(1), 51–65.

Wagner, A., & Plattner, B. (2005). Entropy based worm and anomaly detection in fast ip networks. In *Proceedings of the 14th international workshops on enabling technologies: Infrastructure*

for collaborative enterprise (wetice) (pp. 172–177).

Wagner, D., & Dean, D. (2001). Intrusion detection via static analysis. In *Proceedings of the 22nd symposium on security and privacy (s&p)* (pp. 156–168).

Walker, C., & Pokoski, J. (1985). Residential load shape modelling based on customer behavior. *IEEE Transactions on Power Apparatus and Systems*, *5*(7), 1703–1711.

Wang, W., & Lu, Z. (2013). Cyber security in the smart grid: survey and challenges. *Computer Networks*, *57*(5), 1344–1371.

Wang, Y., Chen, Q., Kang, C., Zhang, M., Wang, K., & Zhao, Y. (2015). Load profiling and its application to demand response: a review. *Tsinghua Science and Technology*, *20*(2), 117–129.

Wang, Y., Gan, D., Sun, M., Zhang, N., Lu, Z., & Kang, C. (2019). Probabilistic individual load forecasting using pinball loss guided lstm. *Applied Energy*, *235*, 10–20.

Wang, Z., & Zheng, G. (2011). Residential appliances identification and monitoring by a nonintrusive method. *IEEE Transactions on Smart Grid*, *3*(1), 80–92.

Wendzel, S., & Keller, J. (2014). Hidden and under control. *Annals of telecommunications*, *69*(7-8), 417–430.

Westerhof, W. (2017, Aug). *Horus scenario - exploiting a weak spot in the power grid.* SMA Vulnerabilities. Retrieved from `https://horusscenario.com/`

Widén, J., Lundh, M., Vassileva, I., Dahlquist, E., Ellegård, K., & Wäckelgård, E. (2009). Constructing load profiles for household electricity and hot water from time-use data–modelling approach and validation. *Energy and Buildings*, *41*(7), 753–768.

Wójcik, A., Łukaszewski, R., Kowalik, R., & Winiecki, W. (2019). Nonintrusive appliance load monitoring: an overview, laboratory test results and research directions. *Sensors*, *19*(16), 3621.

Xiao, Y., Yang, J., Que, H., Li, M. J., & Gao, Q. (2014). Application of wavelet-based clustering approach to load profiling on ami measurements. In *Proceedings of the 4th international conference on electricity distribution (ciced)* (pp. 1537–1540).

Xie, L., Mo, Y., & Sinopoli, B. (2010). False data injection attacks in electricity markets. In *Proceedings of the 1st international conference on smart grid communications (smartgridcomm)* (pp. 226–231).

Xie, M., Han, S., Tian, B., & Parvin, S. (2011). Anomaly detection in wireless sensor networks: a

survey. *Journal of Network and Computer Applications*, *34*(4), 1302–1325.

Yan, Y., Qian, Y., Sharif, H., & Tipper, D. (2012). A survey on cyber security for smart grid communications. *IEEE Communications Surveys & Tutorials*, *14*(4), 998–1010.

Yang, J., Rivard, H., & Zmeureanu, R. (2005). Building energy prediction with adaptive artificial neural networks. In *Proceedings of the 9th international conference on building performance simulation association (ibpsa)* (pp. 15–18).

Yang, S. L., & Shen, C. (2013). A review of electric load classification in smart grid environment. *Renewable and Sustainable Energy Reviews*, *24*, 103–110.

Yao, R., & Steemers, K. (2005). A method of formulating energy load profile for domestic buildings in the uk. *Energy and Buildings*, *37*(6), 663–671.

Yaseen, Y., & Ghita, B. (2017). Peak-to-average reduction by community-based dsm. In *Proceedings of the 5th international conference on smart energy grid engineering (sege)* (pp. 194–199).

Yip, S. C., Tan, W. N., Tan, C., Gan, M. T., & Wong, K. S. (2018). An anomaly detection framework for identifying energy theft and defective meters in smart grids. *Electrical Power & Energy Systems*, *101*, 189–203.

Yohanis, Y. G., Mondol, J. D., Wright, A., & Norton, B. (2008). Real-life energy use in the uk: how occupancy and dwelling characteristics affect domestic electricity use. *Energy and Buildings*, *40*(6), 1053–1059.

Zakaria, Z., & Lo, K. (2009). Two-stage fuzzy clustering approach for load profiling. In *Proceedings of the 44th international universities power engineering conference (upec)* (pp. 1–5).

Zeifman, M. (2012). Disaggregation of home energy display data using probabilistic approach. *IEEE Transactions on Consumer Electronics*, *58*(1).

Zhang, Y., Meratnia, N., & Havinga, P. (2010). Outlier detection techniques for wireless sensor networks: a survey. *IEEE Communications Surveys & Tutorials*, *12*(2), 159–170.

Zhao, H. X., & Magoulès, F. (2012). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, *16*(6), 3586–3592.

Zhou, Y., Zou, H., Arghandeh, R., Gu, W., & Spanos, C. J. (2018). Non-parametric outliers detection in multiple time series a case study: power grid data analysis. In *Proceedings of the 32nd conference on artificial intelligence (aaai)* (pp. 1–8).

Zoha, A., Gluhak, A., Imran, M. A., & Rajasegarar, S. (2012). Non-intrusive load monitoring

approaches for disaggregated energy sensing: a survey. *Sensors*, *12*(12), 16838–16866.